Social Science
Library

MAY 10 1954

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Analysis of a Verbal Conditioning Situation in Terms of Statistical Learning Theory: W. K. ESTES AND J. H. STRAUGHAN 225

Effect of Number of Choices per Unit of a Verbal Maze on Learning and Serial Position Errors: W. J. BROGDEN AND R. E. SCHMIDT 235

Recall as a Function of Number of Classificatory Categories: R. MATHEWS 241

Performance in Instrumental Learning as a Joint Function of Delay of Reinforcement and Time of Deprivation: C. K. RAMOND 248

Overtraining and Transposition in Children: H. W. STEVENSON AND I. ISCOE 251

Effects of Nonreinforced Trials on Discrimination Learning in Preschool Children: G. N. CANTOR AND C. C. SPIKER 256

Response Consistency in Perception and Retention: E. S. GOLLIN AND A. BARON 259

Apparent Spatial Arrangement and Perceived Brightness: J. E. HOCHBERG AND J. BECK ... 263

The Effects of Experimentally Induced Attitudes upon Task Proficiency: R. B. PAYNE AND G. T. HAUTY 267

Transfer of Perceptual Set: G. A. ECKSTRAND AND D. D. WICKENS 274

Judgment of Time as a Function of Serial Position and Stress: J. L. FALK AND D. BINDRA.. 279

Smooth Eye Movements in the Absence of a Moving Visual Stimulus: G. WESTHEIMER AND D. W. CONOVER 283

American Psychological Association

Vol. 47 No. 4 April 1954

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Copyright, 1954, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 47, NO. 4 APRIL, 1954

ANALYSIS OF A VERBAL CONDITIONING SITUATION IN TERMS OF STATISTICAL LEARNING THEORY[1]

W. K. ESTES AND J. H. STRAUGHAN

Indiana University

It is the purpose of this study to investigate the theoretical significance of a rather striking coincidence between an experimental fact and a mathematical fact. The experimental fact has been established in the Humphreys-type "verbal conditioning" situation. In this situation *S* is asked to predict on each of a series of trials whether some designated event, e.g., the flash of a light, will occur; this event, the analogue of the US in a conditioning experiment, is presented in accordance with a predetermined schedule, usually random with some fixed probability. Several recent investigators (**3, 5**) have noted that *S* tends to match his response rate to the rate of occurrence of the predicted event so that if the probability of the latter is, say, .75, the mean response curve for a group of *S*s tends over a series of trials toward an apparently stable final level at which the event is predicted on approximately 75% of the trials. This behavior has seemed puzzling to most investigators since it does not maximize the proportion of successful predictions and thus does not conform to conventional law of effect doctrine. The mathematical fact which will concern us appeared in the course of developing the formal consequences of statistical association theory (**1, 2**); in a simple associative learning situation satisfying certain conditions of symmetry, the theoretical asymptote of response probability turns out to be equal to the probability of reinforcement. The reasoning involved may be sketched briefly as follows.

We consider a situation in which each trial begins with presentation of a signal, or CS; following the signal, one or the other of two reinforcing stimuli, E_1 or E_2, occurs, the probability of E_1 and E_2 during a given series being π and $1-\pi$, respectively. The behaviors available to *S* are categorized into two classes, A_1 and A_2, by experimental criteria. In the verbal conditioning situation, A_1 is a prediction that E_1 will occur, and A_2 a prediction that E_2 will occur on the given trial. We assume that the CS determines a population, S_c, of stimulus elements which is sampled by *S* on each trial, the proportion θ of the elements in this population constituting the effective sample on any one

[1] This research was facilitated by the senior author's tenure as a faculty research fellow of the Social Science Research Council.

trial. The dependence of S's responses upon the stimulating situation is expressed in the theory by defining a conditional relationship such that each element in S_e is conditioned to (tends to evoke) either A_1 or A_2. In order to interpret the formal model in terms of a verbal conditioning experiment, we assume that when an E_1 occurs it evokes from S a response belonging to class A_1, i.e., one which is compatible with the response of predicting E_1 but which interferes with the response of predicting E_2, and that when an E_2 occurs it evokes a response of class A_2. Then on a trial on which E_1 occurs we expect on the basis of association principles (1) that all elements sampled from S_e on the trial will become conditioned to A_1 while on an E_2 trial the sample will be conditioned to A_2. Now if successive trials are sufficiently discrete so that samples from S_e are statistically independent, the probability of an A_1 after Trial n, abbreviated $p(n)$, is defined in the model as the proportion of elements in S_e that are conditioned to A_1, and similarly for the probability of an A_2, $[1-p(n)]$. With these definitions the rule for calculating the change in response probability on an E_1 trial may be stated formally as

$$p(n+1) = (1-\theta)p(n) + \theta \quad (1)$$

and on an E_2 trial as

$$p(n+1) = (1-\theta)p(n). \quad (2)$$

The genesis of these equations will be fairly obvious. The proportion $(1-\theta)$ of stimulus elements is not sampled, and the status of elements that are not sampled on a trial does not change; the proportion θ is sampled and these elements are all conditioned either to A_1 or to A_2 accordingly as an E_1 or an E_2 occurs.[2] Now in a random reinforcement situation, Equation 1 will be applicable on the proportion π of trials and Equation 2 on the proportion $(1-\pi)$; then the average probability of A_1 after Trial $n+1$ will be given by the relation

$$\begin{aligned}\bar{p}(n+1) &= \pi[(1-\theta)\bar{p}(n) + \theta] \\ &\quad + (1-\pi)(1-\theta)\bar{p}(n) \quad (3) \\ &= (1-\theta)\bar{p}(n) + \theta\pi.\end{aligned}$$

If a group of Ss begins an experiment with the value $\bar{p}(0)$, then at the end of Trial 1 we would have

$$\bar{p}(1) = (1-\theta)\bar{p}(0) + \theta\pi,$$

at the end of Trial 2

$$\begin{aligned}\bar{p}(2) &= (1-\theta)[(1-\theta)\bar{p}(0) \\ &\qquad + \theta\pi] + \theta\pi \\ &= \pi - [\pi - \bar{p}(0)](1-\theta)^2,\end{aligned}$$

and so on for successive trials; in general it can be shown by induction that at the end of the nth trial

$$\bar{p}(n) = \pi - [\pi - \bar{p}(0)](1-\theta)^n. \quad (4)$$

Since $(1-\theta)$ must be a fraction between zero and one, it will be seen that Equation 4 must be a negatively accelerated curve running from the initial value $\bar{p}(0)$ to the asymptotic value π.

This outcome of the statistical learning model is rather surprising at first since it makes asymptotic response probability depend solely upon the probability of reinforcement. It seems, however, to be in excellent agreement with the experimental results of Grant, Hake, and Hornseth (3) and Jarvik (5). The question that interests us

[2] Consequently the functions derived in this paper should be expected to apply only to learning situations which are symmetrical in the following sense. To each response class there must correspond a reinforcing condition which, if present on any trial, ensures that a response belonging to the class will terminate the trial. These functions should, for example, be applicable to learning of a simple left-right discrimination with correction; but not to a left-right discrimination without correction, to free responding in the Skinner box, or to Pavlovian conditioning.

now is whether this agreement is to be regarded as a remarkable coincidence or as a confirmation of the theory. We cannot estimate a confidence level for the latter conclusion since the experiments were not conducted specifically to test the theory, and we cannot guarantee that we would be as alert to notice results contrary to the theory which might appear in the literature as we have been in the case of these decidedly positive instances. It has seemed to us that the least objectionable way out of this impasse is to carry out some new experiments, making use of one of the convenient features of a mathematical theory, namely, that if it will generate one testable prediction for a given experimental situation, it can generally be made to yield many more. In the experiment to be reported we have tried to set up a situation similar in essentials to that used by Humphreys, Grant, and others with an experimental design which would permit testing of a variety of consequences of the theory. Each *S* was run through two successive series of 120 trials in an individualized modification of the Humphreys situation with the schedule of π values shown in Table 1. Within the first series we will be able to compare learning rates and asymptotes of groups starting from similar initial values but exposed to different probabilities of reinforcement; within the second series we will be able to compare groups starting at different initial values but exposed to the same probabilities of reinforcement. Comparison of Group I with the other groups over both series will permit evaluation of the stability of learning rate (θ value) from series to series when the π value does or does not change. Series I_A and series II_B will provide a comparison in which initial response probabilities and π values are the same but the amount of preceding reinforcement differs. In order to separate the effect of over-all π value from that of particular orders of event occurrences, each of the three groups indicated in Table 1 has been subdivided into four subgroups of four *S*s each; within a treatment group, say Group I, all subgroups have the same π value but each receives a separate randomly drawn sequence of E_1's and E_2's.

TABLE 1

EXPERIMENTAL DESIGN IN TERMS OF PROBABILITY OF REINFORCEMENT (π VALUE) DURING EACH SERIES

Group	*N*	Trials 1–120 Series A	Trials 121–240 Series B
I	16	.30	.30
II	16	.50	.30
III	16	.85	.30

METHOD

Apparatus.—The experiment was run in a room containing a 2-ft. square signal board and four booths. Upon the signal board were mounted 12 12-v., .25-amp. light bulbs spaced evenly in a circle 18 in. in diameter. The bulbs occupied the half-hour positions of a clock face. Only the top two lights on the board were used as signals in this experiment. The signal board was mounted vertically on a table 40 in. high and was about 5 ft. in front of *S*s' booths.

The booths were made from two 30 × 60 in. tables, 30 in. high, placed end to end but meeting at an angle so that *S*s sitting behind them would be facing almost directly toward the signal board, about 7 ft. in front of *S*s' eyes. Two *S*s sat at each table. The four *S*s were separated from one another by panels 2 ft. high and 32 in. wide. These panels were mounted vertically on the table tops so as to extend 14½ in. beyond the edge of the table between the seated *S*s.

In each booth, 18 in. back from *S*'s edge of the table, was a wooden panel 12 in. high mounted vertically on the table top and extending across the width of the booth. On the side of this panel facing *S* were two reinforcing lights of the same size as those on the signal board but covered by white, translucent lenses. These lights were directly in front of *S*, 4 in. apart and 8 in. above the table top. On the table below each reinforcing light was a telegraph key.

The orders of presentation and the durations of the signal lights and reinforcing lights were

controlled by a modified Esterline-Angus recorder using a punched tape and a system of electrical pick-up brushes. The recorder was placed on the table behind the signal board. Recorder pens which were activated by depression of the telegraph keys in *Ss*' booths were mounted between the brushes. Thus, the presentations of the lights and *Ss*' responses were recorded on the same tape. A panel light was mounted above the Esterline-Angus recorder so that *E*, seated behind the signal board, could watch the operation of brushes and pens during the experiment.

Windows in the experimental room were covered with opaque material and the experiment was run in darkness except for light that came from the apparatus.

Subjects.—The *Ss* were 48 students obtained from beginning lecture courses in psychology during the fall semester of 1952 and assigned at random to experimental groups.

Procedure.—At the beginning of a session, *Ss* were brought into the room, asked to be seated, and read the following instructions:

"Be sure you are seated comfortably; it will be necessary to keep one hand resting lightly beside each of the telegraph keys throughout the experiment and to watch both the large board in the front of the room and the two small lights in your own compartment. Your task in this experiment will be to outguess the experimenter on each trial, or at least as often as you can. The ready signal on each trial will be a flash from the two top lights on the big board. About a second later either the left or the right lamp in your compartment will light for a moment. As soon as the ready signal flashes you are to guess whether the left or the right lamp will light on that trial and indicate your choice by pressing the proper key. If you expect the left lamp to light, press the left key; if you expect the right lamp to light, press the right key; if you are not sure, guess. Be sure to make your choice as soon as the ready signal appears, press the proper key down firmly, then release the key before the ready signal goes off. It is important that you press either the left or the right key, never both, on each trial, and that you make your decision and indicate your choice while the signal light is on.

"Now we will give you four practice trials."

At this point the overhead lights were extinguished and the recorder started. If any obvious mistakes were made by *S* during the four practice trials, they were pointed out by *E*. During the four practice trials the reinforcing lights were always given in the order: E_1, E_1, E_2, E_2. After the practice trials the following instructions were read:

"Are you sure you understand all of the instructions so far? The rest of the trials will have to be run off without any conversation or other interruptions. Please make a choice on every trial even if it seems difficult. Make a guess on the first trial, then try to improve your guesses as you go along and make as many correct choices as possible."

Questions were answered by rereading or paraphrasing the appropriate part of the instructions. If there were any questions about tricks the following additional paragraph was read.

"We have told you everything that will happen. There are no tricks or catches in this experiment. We simply want to see how well you can profit from experience in a rather difficult problem-solving situation while working under time pressure."

The recorder was now started again and the 240 experimental trials were run off in a continuous sequence with no break or other indication to *S* at the transition from Series A to Series B. On each trial, the signal lamps were lighted for approximately 2 sec.; 1 sec. later the appropriate reinforcing light in each *S*'s booth lighted for .8 sec.; then after an interval of .4 sec. the next ready signal appeared; and so on. The high rate of stimulus presentation was used in order to minimize verbalization on the part of *Ss*.

Results and Discussion

Terminal response probabilities.—It will be clear from our discussion of Equation 4 that the predicted asymptote for each series will be the value of π obtaining during the series. We have taken the mean proportion of A_1 responses during the last 40 trials of each series as an estimate of terminal response probability, and these values are summarized for all groups and both series in Table 2.

TABLE 2

Terminal Mean Response Probabilities for Each Series

Group	Series A			Series B		
	$\bar{P}$	π	t	$\bar{P}$	π	t
I	.37	.30	2.35	.28	.30	0.77
II	.48	.50	0.55	.37	.30	2.56
III	.87	.85	0.55	.30	.30	0.05
F	69.31			2.98		

For the first series a simple analysis of variance yields an F significant beyond the .001 level for differences among means. From the within-groups variance estimate we obtain a value for the standard error of a group mean, and this is used in the t test between each group mean and the appropriate theoretical mean. For the second series the between-groups F has a probability between the .05 and .10 levels. In neither series were differences among subgroup means significant at the .05 level.

The interpretation seems straightforward. Group III approximates the theoretical asymptote in both series. Group I falls significantly short of the theoretical asymptote in the first series but approximates it in the second series. Group II falls significantly short of the theoretical asymptote in the second series, but reaches the same probability level as had Group I in the first series. Of the t tests computed for differences between the last two blocks of 20 trials in each series, all yielded probabilities greater than .10 except the t for Series II_B which was significant at the .02 level. Evidently the predictions concerning mean asymptotic values are correct, but the rate of approach to asymptote is faster with Group III than under the other conditions.

According to theory, not only group means, but also individual curves should approach π asymptotically. To obtain evidence as to the tenability of this aspect of the theory we have examined the distributions of individual A_1 response proportions for the last 40 trials of Series III_A, III_B, and I_B. If all individual p values approximate the theoretical asymptotes over these trials, then for each of the series the individual response proportions should cluster around the mean value, π, with an approximately binomial

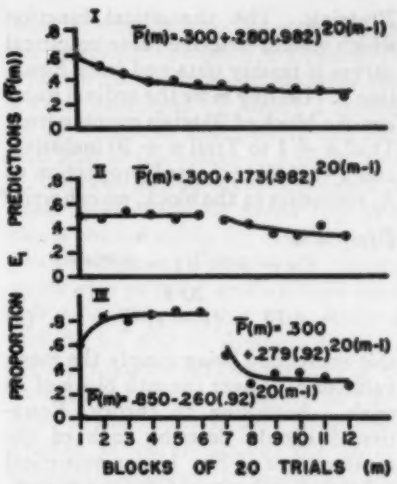

FIG. 1. Empirical and theoretical curves representing mean proportion of E_1 predictions (A_1 responses) per 20-trial block for each series

distribution. Taking the theoretical σ equal to $\sqrt{40\pi(1-\pi)}$, which is actually a slight underestimate of the true value, we find that approximately half of the scores in each series fall within one σ of the theoretical asymptote and only one score in each series deviates by more than three σ. It appears, then, that except for a few widely deviant cases the p values of individual *Ss* approach the theoretical asymptote.

One might raise a question as to just what is meant by the asymptote of an empirical curve in a situation of this kind. Naturally one would not expect the *Ss* to perform at constant rates indefinitely. It does not seem that any sort of breaking point was approached in the present study, however; one subgroup of Group I was run for an additional 60 trials beyond Trial 240 and maintained an average proportion of .304 A_1 responses over these trials.

Mean learning curves.—In Fig. 1 mean data are plotted in terms of the proportion of A_1 responses per block of

20 trials. The theoretical function which should describe these empirical curves is readily obtained from Equation 4. Letting m be the ordinal number of a block of 20 trials running from Trial $n + 1$ to Trial $n + 20$ inclusive, and $\bar{P}(m)$ the expected proportion of A_1 responses in the block, we can write

$$\bar{P}(m) = \pi - \frac{[\pi - \bar{p}(0)](1 - \theta)^{20(m-1)}}{20\,\theta} \times [1 - (1 - \theta)^{20}] \quad (5)$$

this expression being simply the mean value of $p(n)$ over the mth block of 20 trials. According to theory, Equation 5 should describe each of the mean curves of Fig. 1 once numerical values are substituted for the parameters π, $\bar{p}(0)$, and θ; furthermore, the value of θ required should not differ among groups within either series and should be constant from series to series for each group. The values of π are of course fixed by the experimental procedure. The values of $\bar{p}(0)$ in the first series should be in the neighborhood of .50, but for groups of size 16 sampling deviations could be quite large so it will be best to get rid of $\bar{p}(0)$ in favor of $\bar{P}(1)$ which can be measured more accurately. To do this we write Equation 5 for $m = 1$

$$\bar{P}(1) = \pi - \frac{[\pi - \bar{p}(0)]}{20\,\theta}[1 - (1 - \theta)^{20}]$$

then solve for $[\pi - \bar{p}(0)]$

$$[\pi - \bar{p}(0)] = \frac{20\,\theta\,[\pi - \bar{P}(1)]}{1 - (1 - \theta)^{20}}$$

and substitute this result into Equation 5 giving

$$\bar{P}(m) = \pi - [\pi - \bar{P}(1)](1 - \theta)^{20(m-1)}. \quad (6)$$

Observed values of $\bar{P}(1)$ turn out to be .58 and .59 for Series I_A and III_A, respectively. Now we lack only empirical estimates of θ and these can be obtained by a simple statistical procedure. The method we have used is to sum Equation 6 over all values of m, obtaining for K blocks of trials

$$\sum_{m=1}^{K} \pi - [\pi - \bar{P}(1)](1 - \theta)^{20(m-1)} = K\pi - [\pi - \bar{P}(1)]\frac{[1 - (1 - \theta)^{20K}]}{1 - (1 - \theta)^{20}} \quad (7)$$

then equate Equation 7 to the sum of the observed proportions for a given series and solve for θ. For Group I we obtain the estimate $\theta = .018$ and for III_A, $\theta = .08$. Using these parameter values we have computed the theoretical curves for Group I and for the first series of Group III, which may be seen in Fig. 1. In this analysis we find agreement between data and theory in one respect but not in another. The theoretical curves provide reasonably good descriptions of the observed points, especially in the case of Group I, but the θ values for the two groups are by no means equal. The latter finding does not come as a surprise inasmuch as we had found in the previous section that Group I was significantly short of its theoretical asymptote in the first series, while Group III was not.

We did not try to estimate a θ value for the first series of Group II since the empirical curve is virtually horizontal and closely approximates the line $\bar{P}(m) = \pi = .50$. We could proceed to estimate θ values for Series II_B and III_B by the method used above, but it will be of more interest to construct predicted curves for these series without using any additional information from the data. According to the theory, it should be possible to compute those curves from information already at our disposal. The $\bar{p}(0)$ values in the second series should be

TABLE 3

PREDICTED AND OBSERVED MEAN FREQUENCIES OF THE A_1 RESPONSE IN THE SECOND SERIES

Group	Observed	Predicted	t
I	37.19	37.74	0.22
II	46.00	45.94	0.02
III	42.75	42.86	0.04
F	3.16 ($p>.05$)		

the theoretical asymptotes of the first series, or .50 and .85 for Groups II and III, respectively. The only procedural difference between I_A and II_B lies in the number of preceding reinforcements; according to the statistical model, however, this variable will be expected to have no effect except insofar as it leads to a change in $\bar{p}(0)$, so except for sampling error the θ value estimated for Group I should be applicable to II_B. Using .50, .30, and .018 as the values of $\bar{p}(0)$, π, and θ, respectively, we have computed a theoretical curve for Series II_B, and this is plotted in Fig. 1. Similarly, the θ value estimated for Series III_A should apply also to III_B, and we have used this value, .08, together with .30 for π and .85 for $\bar{p}(0)$ to compute the predicted curve for III_B shown in Fig. 1. Considering that no degrees of freedom in the Series B data have been utilized in curve fitting, the correspondence between the theoretical and

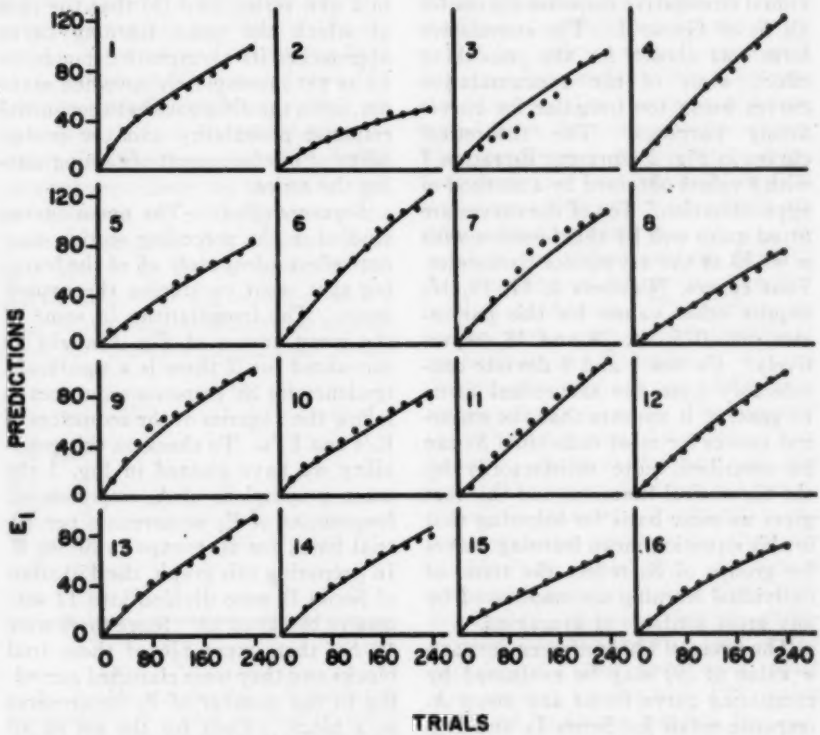

FIG. 2. Empirical and theoretical cumulative response curves for individual *Ss* of Group I

empirical curves does not seem bad. The reason for some of the irregularities will be brought out in the next section. A statistical test of one aspect of the correspondence can be obtained by calculating for each theoretical curve a predicted mean total of A_1 responses in the second series, by means of Equation 5, and comparing these values with the observed mean totals. This has been done and the comparison is given in Table 3. The t values for differences between observed and theoretical values seem satisfactorily low.

In order to give an idea of the extent to which the behavior of individual *Ss* conforms to the theoretical function, we have plotted in Fig. 2 the individual cumulative response curves for all *Ss* of Group I. The cumulative form was chosen for the smoothing effect, some of the noncumulative curves being too irregular for curve-fitting purposes. The theoretical curves in Fig. 2 represent Equation 7 with θ values obtained by a method of approximation. Ten of the curves are fitted quite well by this function with $\pi = .30$ as the asymptote parameter. Four curves, Numbers 2, 11, 15, 16, require other values for this parameter, viz., .075, .45, .24, and .18, respectively. Curves 3 and 4 deviate considerably from the theoretical form. In general, it appears that the empirical curves for most individual *Ss* can be described quite satisfactorily by the theoretical function, and this fact gives us some basis for inferring that in this situation mean learning curves for groups of *Ss* reflect the trend of individual learning uncomplicated by any gross artifacts of averaging.

The effect of 120 reinforcements at a π value of .50 may be evaluated by comparing curve forms and mean A_1 response totals for Series I_A and II_B. We find that the reinforcements lead to no increase in resistance to change. Slopes of the two curves are very similar and the response totals do not differ significantly. This result is in line with predictions from the statistical model, but a little surprising, perhaps, from the viewpoint of Thorndikian or Hullian reinforcement theory since partial reinforcement has generally (6) been held to increase resistance to extinction in this situation.

The conclusions from our study of the mean learning curves would seem to be (*a*) that under some circumstances at least it is possible to evaluate theoretical parameters from the data of one series of learning trials and then to predict the course of learning in a new series; and (*b*) that the rate at which the mean learning curve approaches its asymptote depends, in an as yet incompletely specified manner, upon the difference between initial response probability and the probability of reinforcement obtaining during the series.

Sequence effects.—The mean curves studied in the preceding section may not reflect adequately all of the learning that went on during the experiment. The irregularities in some of the mean curves of Fig. 1 might be accounted for if there is a significant tendency for *Ss*' response sequences to follow the vagaries of the sequences of E_1's and E_2's. To check on this possibility we have plotted in Fig. 3 the mean proportions of A_1 responses vs. frequencies of E_1 occurrences per 10-trial block for all groups in Series B. In preparing this graph, the 120 trials of Series B were divided into 12 successive blocks of 10. Since there were 48 *Ss*, there were 576 of these trial blocks and they were classified according to the number of E_1 occurrences in a block. Then for the set of all blocks in which no E_1's occurred, the

mean proportion of A_1 responses was computed and entered as the first point in Fig. 3, and so on for the remaining points. It seems clear that *Ss* were responding to the particular sequences of E_1's and E_2's, and not simply to the over-all rate. Corresponding graphs for the three groups in the first series had somewhat shallower slopes; they have not been reported since some of the individual points were based on too few cases to be reliable and the groups could not be averaged together in the first series owing to the different π values.

In order to deal statistically with this apparent dependence of response tendency upon the density of E_1 occurrences in the immediately preceding sequence, we have computed for each series the average probability, $\bar{p}_{A_1|E_1}$, that an A_1 occurs on Trial n given that an E_1 occurs on Trial $n-1$ and the average probability, $\bar{p}_{A_1|E_2}$, that an A_1 occurs on Trial n given that an E_2 occurs on Trial $n-1$. The difference between these two quantities can be shown to be proportional to the point correlation (7) between $A(n)$ and $E(n-1)$ for a given series. Furthermore our Equations 1 and 2 may be regarded as theoretical expressions for the two conditional probabilities, $\bar{p}_{A_1|E_1}$ and $\bar{p}_{A_1|E_2}$, respectively, and it will be seen that if these expressions are averaged over all values of n in a series and the second subtracted from the first, the difference is equal to the parameter θ, i.e.,

$$(1-\theta)\bar{p}(n) + \theta - (1-\theta)\bar{p}(n) = \theta.$$

Thus from the statistical model we must predict that the difference between empirical estimates of these conditional probabilities for any series should be positive and, if successive trials are independent, this difference should be equal to the value of θ estimated from the mean response curve. The conditional probabilities have been computed from the data for each *S* and mean differences by groups are summarized in Table 4.

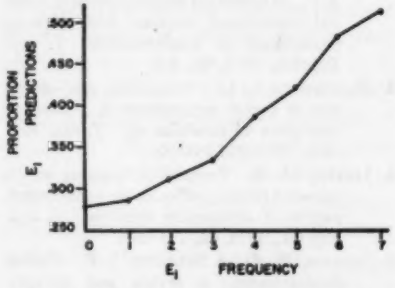

FIG. 3. Mean proportion of E_1 predictions (A_1 responses) in a block of ten trials plotted against the actual number of E_1 occurrences in the block; data averaged for all groups in Series B

TABLE 4

MEAN DIFFERENCES BETWEEN OBSERVED VALUES OF $\bar{p}_{A_1|E_1}$ AND $\bar{p}_{A_1|E_2}$ FOR EACH SERIES

Series	Group I	Group II	Group III
A	.128	.199	.153
B	.214	.294	.231

All of the differences are positive and significant at better than the .001 level of confidence. The differences among group means are insignificant for both series (*F*'s equal to .45 and .73, respectively) as are differences among subgroup means. The increases from the first series to the second are, however, significant beyond the .005 level. The latter effect was not anticipated on theoretical grounds; the most plausible explanation that has occurred to us is that alternation tendencies associated with previously established guessing habits extinguished during the early part of the experiment. This hypothesis would also account for the high $\bar{P}(1)$ value observed for Group I in Fig. 1.

Although all of the quantities in Table 4 are positive and apparently

independent of π, as required by the theory, the numerical values are all larger than the θ estimates obtained from mean response curves. The most straightforward interpretation of this disparity would be that, owing to the short intertrial interval, successive trials are not independent in the sense required by the theoretical model. Nonindependence would have at least two immediate consequences in so far as the present experiment is concerned. First, stimulus samples drawn on successive trials would overlap, and the learning that occurred on one trial would affect behavior on the next to a greater extent than random sampling would allow for, thus increasing $\bar{p}_{A_1E_1}$, and decreasing $\bar{p}_{A_1E_2}$. Second, the reinforcing stimulus of one trial, E_1 or E_2, would be part of the stimulus complex effective at the beginning of the next trial. If this interpretation is correct, then more widely spaced trials should result in better agreement between the alternative estimates of θ and also in reduction of the dependence of mean learning rate upon probability of reinforcement.

Summary

Learning rates, asymptotic behavior, and sequential properties of response in a verbal conditioning situation were studied in relation to predictions from statistical learning theory.

Forty-eight college students were run in an individualized modification of the "verbal conditioning" experiment originated by Humphreys (4). Each trial consisted in presentation of a signal followed by a left-hand or right-hand "reinforcing" light; *S* operated an appropriate key to indicate his prediction as to which light would appear on each trial. For each *S* one of the lights, selected randomly, was designated as E_1, the other as E_2. On the first series of 120 trials, E_1 occurred with probability .30, .50, and .85 for Groups I, II, and III, respectively. On the second 120 trials, E_1 occurred with probability .30 for all groups.

Theoretical predictions were that mean probability of predicting E_1 should tend asymptotically to the actual probability of E_1, both during original learning and following a shift in probability of reinforcement; and that response probabilities should change in accordance with exponential functions, learning rates (as measured by slope parameters) being independent of both initial condition and probability of reinforcement.

The statistical criterion for approach to theoretical asymptote was met by Group I by the end of the second series and by Group III in both first and second series. In the second series, Group II was short of theoretical asymptote but reached the same response probability as had Group I during the first series.

Learning rates were virtually identical for Group I, first series, and Group II, second series, indicating that resistance of response probability to change is not altered by 50% random reinforcement in this situation. Learning rates differed significantly among groups within both series. In general, learning rate was directly related to difference between initial response probability and probability of reinforcement during a series. It was suggested that this relationship may depend upon temporal massing of trials. Not only group means, but individual learning curves could be described satisfactorily by theoretical functions.

No tendency was observed for *Ss* to respond to a series as a whole. On the contrary, sensitivity to effects of individual reinforcements and nonreinforcements (E_1 and E_2 occurrences) increased significantly as a function of trials.

References

1. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
2. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
3. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
4. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
5. Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.
6. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
7. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.

(Received July 10, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

EFFECT OF NUMBER OF CHOICES PER UNIT OF A VERBAL MAZE ON LEARNING AND SERIAL POSITION ERRORS[1]

W. J. BROGDEN AND ROBERT E. SCHMIDT

University of Wisconsin

The experiment to be reported in this paper was designed as an investigation on the acquisition of spatial orientation. The *S*s were asked to assume movement within a spatial maze as they responded with words that represent movement in space. Mazes of 16 units in length, with 2, 3, 4, 6, or 8 alternate choices per unit were used. The mazes will differ in spatial complexity only if *S* is able to assume movement in space as he makes the required responses. Since the *S*s failed in achieving this attitude, the task becomes one of nonspatial serial verbal learning.

The experiment, then, provides for a determination of the relation between acquisition of a verbal maze and the number of alternate choices per unit. This variable is representative of the dimension of amount of discovery as defined by Melton (3). With other conditions equal, difficulty of acquisition should increase with increase in amount of discovery required. Riley (4) has demonstrated this to be so for the acquisition of a list of eight paired associates where the correct response was discovered among 1, 2, or 4 alternatives. The present experiment extends the range of amount of discovery and provides information on the relation of this variable to difficulty of acquisition with parameters of learning different from those present in the study by Riley.

[1] This research was supported in part by a grant from the University Research Committee from funds provided by the Wisconsin Alumni Research Foundation.

Procedure

Twenty *S*s in each of five groups practiced a 16-unit verbal maze until the criterial performance of one perfect repetition was attained. The number of choice points per unit was 2, 3, 4, 6, and 8, respectively, for the five groups. The *S*s were members of the classes in elementary psychology at the University of Wisconsin who volunteered to serve in an experiment on learning a mental maze. Of the total 100 *S*s, 76 were women and 24 were men. The *N* of 20 for each of the five groups was filled simultaneously by random assignment of *S*s to each group or subgroup.

Conditions.—Each unit of Maze 1, the two-choice maze learned by Group I, involved a choice of right (R) or left (L). For Maze 2, the three choices were R, L, and up (U), or R, L, and down (D). Half of the *S*s in Group II (Subgroup II*a*) was given the choices R, L, and U, and the other half (Subgroup II*b*) was given the choices R, L, and D. The pattern for Maze 2 was identical in these two cases except for the direct interchange of the choices, U and D. In Maze 3, the *S*s of Group III had the four choices of R, L, U, and D for each unit. Group IV was divided into two equal subgroups, of which IV*a* had the choices R, L, U, D, right-up (RU), and left-up (LU); Subgroup IV*b* was given the choices R, L, U, D, right-down (RD), and left-down (LD). The pattern of Maze 4 was identical for these two subgroups except for the direct interchange of the choices RU and RD with the choices LU and LD, respectively. The eight choices per unit in Maze 5 were R, L, U, D, RU, LU, RD, LD. Table 1 presents the pattern of correct choice for each of the mazes. The sequences of correct choice were obtained by use of a table of random numbers with the restrictions that the first choice be R and that there be an equal number of each kind of choice. Since the latter condition could not be met completely for Maze 2 (three-choice) and Maze 4 (six-choice), equality for kind of choice was held to five for Maze 2 and two for Maze 4; the correct choices for the remaining units were filled at random with the requirement that a given kind of choice not appear more than once.

Instructions.—Models of one unit of each of the five mazes were constructed from Tinker

TABLE 1

MAZE PATTERNS

Group	Maze Number	Number of Choices	Units of Maze															
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
I	1	2	R	L	L	R	R	R	L	R	R	L	L	L	R	L	R	L
II*a*	2*a*	3	R	U	L	L	R	R	U	U	L	R	L	U	L	L	R	U
II*b*	2*b*	3	R	D	L	L	R	R	D	D	L	R	L	D	L	L	R	D
III	3	4	R	D	R	D	L	U	R	R	L	D	U	L	L	U	U	D
IV*a*	4*a*	6	R	L	RU	R	U	D	D	RU	RU	D	LU	L	L	R	U	LU
IV*b*	4*b*	6	R	L	RD	R	U	D	D	RD	RD	D	LD	L	L	R	U	LD
V	5	8	R	U	L	U	D	D	R	RD	RD	LU	LU	L	LD	LD	RU	RU

Toy parts. A shaft inserted in the center hole of a circular hub represented one unit of distance forward. The hub represented the choice point, and shafts inserted around the circumference of the hub represented the alternative choices. The angular location of shafts and their designation in terms of verbal response were as follows: 0° = U; 45° = RU; 90° = R; 135° = RD; 180° = D; 225° = LD; 270° = L; 315° = LU. Cardboard tabs with the appropriate words printed on them were inserted at the end of each of the shafts attached to the hub circumference. The *E* used the models in giving instructions to *S*.

The *S* and *E* sat in chairs located at a table in the experimental room with *S* always seated to the right of *E*. The instructions were read by *E*, who also used the appropriate model of the maze unit and pointed to its parts when pertinent to the instructions. The instructions differed for the five maze patterns only in terms of the number and kinds of choices. They asked *S* to assume presence at a given point with the intention of proceeding in a forward direction, as rapidly as possible, to some other point. Mention was made of the possible choices and the model of one unit was demonstrated to show forward movement along the horizontal shaft to the hub or choice point. The alternate choices were again named and their spatial locations indicated on the model. Behavior for both correct and incorrect responses was demonstrated. If the *E* said "forward" following a response, *S* was to enter the next unit and make response from the alternate choices. The end of the maze or trial was indicated by the word "stop" spoken by *E*. A blindfold, composed of a pair of goggles behind each lens of which was placed sterile, absorbent cotton, was adjusted over the eyes. At this point, the instructions were presented again, but the model was not used. Any questions relating to clarification of the instructions were answered and the first trial was then started.

Measures.—The *E* used two stop watches, one to measure the time consumed by each trial, and the other to maintain a 30-sec. interval between each trial. Data sheets were prepared for each maze pattern with the letter symbols of the true pathway printed thereon. The *E* recorded each error by letter symbol and in terms of maze locus and sequence of error.

Terminal questions.—When *S* had completed the criterial trial, the blindfold was removed, and he was asked to answer the following questions: (*a*) How many units were there in the maze? (*b*) How many times was each response used in the maze pattern? (*c*) Can you now give the correct maze pattern? (*d*) Were you able to visualize the correct maze pathway as you were learning the maze? The *S* was cautioned not to attempt answer of *a* and *b* by a count in review of the true pathway. The answers to these questions were recorded by *E*.

RESULTS

The standard measures of maze learning, trials to the criterion, total errors to the criterion, and total time to the criterion are available as measures of acquisition. The number of trials to the criterion, the total errors to the criterion, and the total time in seconds to the criterion were tabulated for each *S* in each of the five groups. An initial examination of these data was made for Subgroups II*a* and II*b* and for Subgroups IV*a* and IV*b*. The mazes for the *a* and *b* subgroups differed from each other only in terms of one or two response words, and not in terms of pattern or in terms of number of alternate choices per unit. Since no significant differences were found, the data for the *a* and *b* sub-

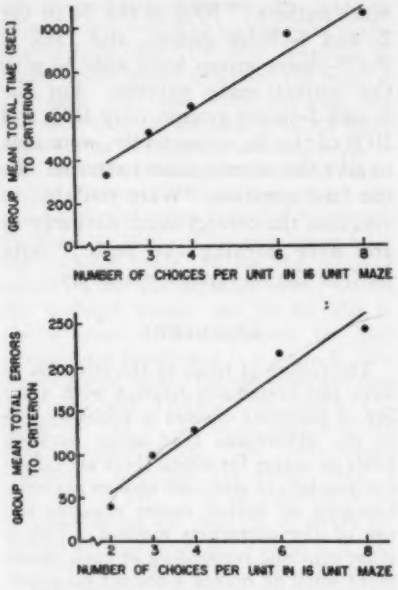

FIG. 1. Relation between acquisition of 16-unit verbal maze and number of alternate choices per unit

groups were pooled and are presented only for the over-all groups, II and IV. An analysis of variance was performed on the data for each of the three measures. The F ratio for the between- and within-groups variances is 2.68 for trials to the criterion, 33.33 for total errors to the criterion, and 22.13 for the total time to the criterion. The value of F at the 5% level of confidence is 2.52. The group means for trials to the criterion are 10.7, 13.3, 12.8, 13.5, and 9.8, respectively, for the five groups. No systematic relation exists between trials to the criterion and number of alternate choices per choice point. The group means for total errors and for total time, on the other hand, show a progressive increase as the number of alternate choices per unit increases. The relations between the means for errors and time to number of alternate choices per choice point were tested for deviation from linear regression (6). In neither case was there significant deviation from linear regression. Therefore, regression equations of the straight line function, $y = mx + b$, were fitted by the method of least squares to the empirical data. These equations are: $y = 115.4x + 157.4$ for time and $y = 34.1x - 12.8$ for errors. Figure 1 presents plots of these curves and of the means for time and errors.

The present experiment also provides information about the relation between the elimination of errors as a function of serial position within the maze and the number of alternate choices per maze unit. Total errors to the criterion were tabulated for each S for each of the 16 units in the maze and an analysis of variance was performed on these data. The F for groups was 216.82, for units of the maze was 34.02, and for the interaction of groups by units was 6.41. Each of these F ratios is significant at better than the 5% level of confidence. The significant F for groups indicates that more errors are made by Ss as

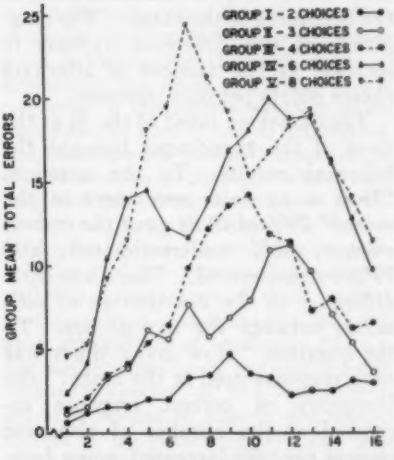

FIG. 2. Serial position error curves for verbal mazes with different numbers of choices per maze unit

the number of choice points per maze unit increases. The one for units of the maze indicates that frequency of error varies with serial position of maze unit. The significant interaction term shows that the serial position error curve differs as a function of the number of alternate choices per maze unit. These results are shown in Fig. 2 where group mean frequency of total error is plotted as a function of unit position within the maze for each group. Each of the five curves was smoothed by taking a rolling average of "threes." The extent of bowing of the serial position curve increases as the number of alternate choices per unit increases. The maximum frequency of error occurs beyond the middle of the maze for all groups except Group V. The curve for this group is bimodal with a maximum frequency of error at Unit 7, but with a secondary peak at Unit 5. Another feature of these curves is their asymmetry with respect to the initial and terminal units of the maze. For all groups there is a greater frequency of error at the end of the maze than at the beginning of the maze. The magnitude of this difference appears to increase as the number of alternate choice points per unit increases.

The questions asked of the *S*s at the close of the experiment brought the following results. To the question, "How many units were there in the maze?" 29% of all *S*s gave the correct answer; 52% underestimated; and 19% overestimated. There was little difference in the distribution of estimates between the five groups. To the question, "How many times was each response used in the maze?" the frequency of correct response increased as the number of alternate choices per unit increased, going from 20% for the 2-choice group to 90% for the 8-choice group. To the question, "Can you now give the correct maze pattern?" 70% of the *S*s in the 2- and 4-choice groups, and 75% in the 8-choice group were able to give the correct maze pattern. For the 3- and 5-choice groups, only 40% and 30% of the *S*s, respectively, were able to give the correct maze patterns. To the final question, "Were you able to visualize the correct maze pathway as you were learning the maze?" only 18% of the *S*s replied positively.

Discussion

The failure of trials to the criterion to have any systematic relation with number of alternate choices is probably due to the differences that occur between trials on mazes for which there are different numbers of alternate choices per unit. Learning of spatial mazes requires the use of the correction method. The *S* must continue responding at each choice point until he makes a correct response. Only then is he permitted to proceed to the next choice point. Thus, trials on mazes with different numbers of alternate choices are not comparable. If the noncorrection method were used, trials would be an appropriate measure of acquisition where number of alternate choices is the independent variable. In the study by Riley (4), the noncorrection method was used and trials to the criterion show a progressive increase from 1 to 2 to 4 alternate choices.

In the present study, the data for total errors and total time to the criterion show a progressive increase with increase in the number of alternate choices per maze unit. The two functions are essentially the same. What may be said about the one, may also be said of the other. With the 2-choice maze as the base, difficulty of acquisition becomes disproportionately less difficult with increase in the number of alternate choices. Increase in the number of alternate choices per choice point is perhaps comparable to increase in the length of list, quantity of material, or length of a maze. Studies of this kind of variable, however, show a disproportionate increase in difficulty of acquisition with its increase. This result

has been obtained with verbal learning where length of list or quantity of material is increased (1), with learning of finger mazes where length of the maze is increased (5), and with learning the Peterson rational learning problem when the number of letters is increased (2).

A similar relation exists between number of alternate choices per maze unit and number of errors expected by chance on the first trial. This number is 8 for the 2-choice maze, 16 for the 3-choice maze, 24 for the 4-choice maze, 40 for the 6-choice maze, and 56 for the 8-choice maze. The equation for this straight line function is $y = 8x - 8$. The actual group means for errors on the first trial are 7.00, 15.25, 22.55, 39.25, and 63.60 for the 2-, 3-, 4-, 6-, and 8-choice mazes, respectively. The least-square regression equation for these data is $y = 9.3x - 131$. One assumption made in determining the number of errors expected by chance on the first trial is that no error will be repeated for any given choice point. This is a reasonable assumption since *S* was informed of the number and kind of responses he was to make before practice on the maze was started. Examination of the data reveals no repetition errors for *S*s who learned the 2-, 3-, or 4-choice mazes. Two repetitive errors were made in the 6-choice maze and 95 repetitions were made on the 8-choice maze. If the repetitive errors are subtracted from the total errors on the first trial for each *S* in the 6- and 8-choice maze groups, the means are 39.15 and 58.85, respectively. The regression equation for the adjusted data is then $y = 8.6x - 10.9$. The adjusted data are in close agreement with chance expectancy of errors on the first trial. The difference between these functions and that for all of the data represents a difference in relative maze difficulty that, on the one hand, shows a disproportionate increase in difficulty as number of alternate choices increases and, on the other, shows an increase in difficulty as number of alternate choices increases that decreases disproportionately. Since the former function is included in the latter, it follows that relative difficulty between the mazes decreases as practice progresses. Amount of discovery makes its greatest contribution to performance on the first trial; thereafter, its effect declines with further practice.

There is a characteristic of the independent variable other than number of alternate choices per choice point that may be important in producing the decrease in relative difficulty of the mazes as learning progresses. When maze length is held constant as the number of alternate choices per choice point is increased, the number of like responses in the correct maze pattern must decrease. For the 2-choice maze there are 8 R and 8 L responses in the true pathway; for the 3-choice maze, the correct path contains 5+ of each of the three responses; for the 4-choice maze, the number of like responses is 4; for the 6-choice maze, it is 2.66; and for the 8-choice maze it is 2. A decrease in the number of like responses in the correct pathway may lessen difficulty in learning through an increase of discrimination between the 16 units of the maze and a decrease in interference effects between the units. If this factor exists, it can affect learning only after some degree of acquisition of the maze pattern has occurred. Its effect presumably would be greatest in the later stages of acquisition. Upon this theoretical basis, therefore, the primary effect of number of alternate choices per choice point is assigned to the early stages of acquisition, the effect of number of like responses in the correct pathway is assigned to the later stages of acquisition, and the disproportionate decrease in increments of maze difficulty of the total time and total error functions is attributed primarily to the effect of the number of like responses in the correct maze path.

Number of like responses in the correct pathway is an aspect of maze pattern that may have a considerable influence upon the relation of acquisition to number of alternate choices per choice point. Although number of like responses is inversely related to number of alternate choices in the present experiment, it is possible to make them independent. Experimental variation of number of like responses could then be applied to all possible choice responses of the maze and

to possible choice responses in successive units of the maze or blocks of units, as well as to the correct responses. Consideration of these dimensions would also require that consideration be given to the sequence of the correct responses. In the present experiment, maze difficulty is in part a function of the particular sequence of correct responses selected at random. Manipulation of each of these variables might be expected to affect acquisition and thereby alter relative maze difficulty and the functional relation of acquisition to number of alternate choices per choice point.

Note should be made also of the possible effect of further increase in the magnitude of the independent variable upon relative difficulty of maze acquisition. From a theoretical point of view, there is an upper limit to the amount of material, the length of a maze, or the number of alternate choices per choice point of a maze that can be learned as a unit of serially presented material by the whole method. With the kind of verbal maze used in the present experiment, it would be necessary to increase the number of units (or length of the verbal maze) in order to be able to increase the number of alternate choices per unit. If this were done to a sufficient extent, a point of inflection should be found beyond which there would be a nonlinear acceleration of the function to the value of the independent variable at which learning to the criterion will not occur.

The relations between the form of the serial position error curve and number of alternate choices per choice point cannot be given quantitative expression, even though the increase in extent of bowing is clearly evident. However, it should be noted that the characteristics of these curves are comparable to each other and to those for serial position error curves based on learning data for serial verbal material and spatial mazes.

Summary

Five groups of *Ss* learned a 16-unit verbal maze for which the number of alternate choices per unit were 2, 3, 4, 6, or 8. Rate of acquisition measured in terms of mean total time and mean total errors to the criterion of one perfect repetition increased with increase in the number of alternate choices per choice point. Analysis of variance indicates a significant effect of the independent variable on both measures of acquisition. Tests of deviation from linear regression reveal no deviation. Therefore, straight line regression equations were fitted to the data by the method of least squares. These functions are $y = 34.1x - 12.8$ for errors and $y = 115.4x + 157.4$ for time in seconds.

Theoretical discussion of these functions suggests that in the determination of maze difficulty the effect of number of alternate choices per choice point occurs early in acquisition; the effect of number of like responses in the correct pathway occurs later in acquisition; and the number of like responses is primarily responsible for the disproportionate decrease in the increments of maze difficulty.

Serial position error curves were constructed from the data for each group, and extent of bowing of these curves was found to increase with increase in the number of alternate choices per choice point. An analysis of variance shows this relation is significant statistically. Other characteristics of the serial position error curves are comparable to those obtained in studies of serial verbal learning and maze learning.

References

1. Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 613–689.
2. McGeoch, J. A., & Oberschelp, V. J. The influence of length of problem and of transfer upon rational learning and its retention. *J. gen. Psychol.*, 1930, **4**, 154–168.
3. Melton, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. (Rev. Ed.) New York: Macmillan, 1950. Pp. 668–690.
4. Riley, D. A. Rote learning as a function of distribution of practice and the complexity of the situation. *J. exp. Psychol.*, 1952, **43**, 88–95.
5. Scott, T. C., & Henninger, L. L. The relation between length and difficulty in motor learning: a comparison with verbal learning. *J. exp. Psychol.*, 1933, **16**, 657–678.
6. Snedecor, G. W. *Statistical methods*. (4th Ed.) Ames, Ia.: Iowa State Coll. Press, 1948.

(Received June 12, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

RECALL AS A FUNCTION OF NUMBER OF CLASSIFICATORY CATEGORIES[1]

RAVENNA MATHEWS

University of California[2]

Some psychologists are interested in the organism's processing of information—how it is received or selected, organized, stored, made available when needed. *Categories* are important in the processing of information because they establish equivalences in our Heraclitian world, making it possible for the organism to profit by experience and react at a better-than-chance level with the environment (**3, 8, 12**).

It seems evident that the process of categorizing is itself one determinant of which previous experiences are brought to bear on subsequent occasions. That the use of a particular category influences the content of the information that is recalled has been demonstrated in many ways (**1, 4**). However, the influences of formal characteristics of categories, such as how many there are, their distinctness, exclusiveness, "ecological validity," generality, etc., have been studied very little.

The problem of the present study is the apparently simple one of how mere *number* of categories affects the amount of information that can be recalled. A further problem is how such effects can be explained in terms of more molecular psychological concepts concerning cognitive processes.

An illustration may clarify the presentation of the issue. We have found it harder to remember the names of our students when the class is either all male or all female than when the class is mixed—when there are two categories. What if there were classes with 3 sexes, or 4, or 20? Would it be easier and easier to remember names then? Of course, certain conditions must be specified. Let us assume that the categories are distinct, that *S* will have no problem in remembering what the categories are, that an equal number of cases will fall into each category, and that each case falls into its category unambiguously. Also though the above example suggests the method of stimulus-response pairs, let us assume the method of retained members.

An experiment designed to test the effect of the number of categories on recall becomes essentially another study of the effect of similarity on recall. This is so because items do not belong to the same category unless they are in some respect similar. Therefore, with a total number of items that is fixed, increasing the number of categories means decreasing the number of similar items. Suppose there is one group of *S*s who place 24 items into two categories and another group who place 24 items into six categories. Later they are asked to recall the items that they classified. Which group will recall more?

Although Gestalt and S–R memory theorists would analyze this situation somewhat differently, the prediction as to outcome would probably be the same. If items that belong to the same category are similar, one might say that items within a category form a "trace aggregate" from which individual items will be increasingly difficult to extricate as the size of the aggregate increases. Or one might say that the category name is attached to each of the items within it (constant stimulus, different responses),

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of California. The writer is indebted to Dr. Warner Brown, thesis director, and Dr. Leo Postman.

[2] Now at Smith College.

and as the number of items within a category increases, the tendency to recall any one of the items suffers greater interference from the tendencies to give all the other items at the same time.

Our data do not furnish any test of the first formulation. The second is derived from Gibson (5), although she did not attempt to apply these concepts to the recall situation set up by the method of retained members, since it is not subject to detailed stimulus-response analysis. Also, in the case of stimulus-response pairs, only one of the presumably competing responses is correct for each stimulus; in the case of retained members, all of the competing responses are "correct." Nevertheless, others have assumed that an analysis in terms of intraserial interference is applicable to this situation (6, 9, 10).

Although perhaps oversimplified, this analysis is the basis for the hypothesis that increasing the number of categories will increase the number of items that can be recalled.

Method

Because of difficulties in designing one set of material which was equally well classifiable into 2, 3, or 6 categories, the alternative plan was adopted of constructing several lists which were equivalent in familiarity but composed of items only in part the same. The categorizing scheme to be used by *S* was imposed upon him not only by the instructions but also by the selection of items on the list he received.

The general procedure was to give *S* a list of names of famous people, each of which he was to classify as belonging to a certain category. After a 10-min. interval of interpolated activity, he was instructed to write as many of the names as he could remember.

Subjects.—The *S*s were beginning psychology students attending weekly section meetings in groups of 10–25 at the University of California. The standardization of the categories was based on 150 *S*s, and 238 *S*s served in the experiment proper.

The categories and construction of the lists.—The categories were: Artist, Athlete, Conqueror, Musician, Poet, and Scientist. Lists of 24 names belonging to 2, 3, or 6 of these categories were constructed. A 2-category list had 12 items in each category, a 3-category list had 8 items in each category, and a 6-category list 4 items in each category.

The design of the experiment called for sets of names which were equally familiar, whether they were to be classified into 2, 3, or 6 categories. The procedure for obtaining equated lists was as follows. Subjects from the same population as that to be tested in the experiment proper were given 5 min. in which to think of as many members of a given category, e.g., Artists, as they could. At the end of each minute they were told to draw a line beneath the last name they had written. This technique was taken from Bousfield and Sedgewick (2). When the first 5-min. interval was finished, in most cases *S*s were given a second category. No group was given more than two categories. Between 45 and 50 *S*s contributed names for each of the six categories.

The *S*s were told what was to be meant by the category names. Conquerors were to include noncontemporary military or naval figures; Musicians were to be restricted to serious composers of music; Athletes should include great professionals only, etc. The restrictions were *E*'s intuitive attempts to establish categories of comparable size.

After the names were collected, master lists which gave the percentage of *S*s mentioning each recurring name were prepared for all the categories. From these master lists, the experimental series were built, the attempt being to construct lists in which the names were of comparable familiarity.

Equating lists.—The criteria that were set up for equating the lists were (*a*) that the percentage of mention of names in each list should average the same; (*b*) that the percentage of mention of names in each category within a list should average the same; and (*c*) that in each list the number of names mentioned by high, medium, and low percentages of *S*s should be the same. The lists show some deviations from these criteria. Nevertheless, there is evidence that the equating procedure was successful with respect to selection of names within categories.

Since the 12 names of the 2-category group for any one category included all the names belonging to that category which were on both the 3- and 6-category lists, one can look to see whether the four Musicians, for example, that were on both the 6-category list and the 2-category list accounted for about 33% of the total number of Musicians recalled by *S*s in the 2-category group. If these names accounted for more than 33%, it might be that the 6-category group had been given the advantage of especially familiar names, in spite of the efforts to control for familiarity, and would suggest that the con-

TABLE 1

PERCENTAGE OF TOTAL RECALL BY 2-CATEGORY GROUP CONTRIBUTED BY NAMES ON 3- AND 6-CATEGORY LISTS

Category	3-Category List (Expected: 67%)	6-Category List (Expected: 33%)
Artist	69	33
Athlete	66	29
Conqueror	60	35
Musician	68	33
Poet	75	44
Scientist	65	29
Mean	67	34

trol procedure was inadequate. Similarly, the eight Musicians common to the 3-category list and the 2-category list should account for about 67% of the total number of Musicians recalled by the 2-category group. The same should hold for the other categories.

Table 1 shows that this is the case. Only for Poets was there a sizable (positive) variation from the expected percentages. (One might expect to find that the 2-category group would show a disadvantage in its recall of Poets. Instead, later evidence leads us to the interpretation that facilitation effects among the larger number of Poets of the 2-category group were responsible for a particularly high level of recall by that group of some names on the list.)

One further problem in the equating procedure remained. Although certain Musicians and certain Poets, for example, may be mentioned equally often, one does not know whether Musicians or Poets are in general easier to recall. Therefore, 2- and 3-category lists were made using all of the six classes of names. For the 2-category lists, 7 of the 15 possible combinations of six categories taken two at a time were used. For the 3-category lists, 3 combinations of six categories taken three at a time were used.

Table 2 gives the combinations of categories that were used, the average percentage of mention of the names on each list, and the number of *S*s who received each list. Table 3 gives for illustrative purposes the names that were on the 6-category list.[3]

Procedure.—The *S*s received booklets with instructions on the first page. The instructions given to one of the 3-category groups were as follows:

"Your first task will be to identify 24 people as belonging to one of three categories: Poet, Athlete, or Artist. These people will be listed, with a blank preceding each name. If you recognize a name to be that of a Poet, put a 'P' in the blank. Put an 'Ar' if the person is an Artist and an 'Ath' if the person is an Athlete. Be sure that you use these symbols and no others.

"If you do not recognize a name, or if you recognize it but have no idea what category it belongs in, leave the blank empty. Under such circumstances it will not be to your advantage to guess.

"Later you will be asked to give as many of these names as you can. However, you are not to try to memorize the list. Spend only as much time on it as is necessary to identify accurately the names you know. As soon as you have finished, tear off the list and hold it up. The experimenter will take it, and you will proceed to answer some questions on Maladjustment."

The list of names was on the back page of the booklet. With respect to category, the names were given in jumbled order. After *S* gave the back page to *E*, he turned to the second page of the booklet and answered questions on the material that had been assigned for discussion in the quiz section that day. Since the last *S* turned in

TABLE 2

SUMMARY OF COMBINATIONS OF CATEGORIES USED IN 2-, 3-, AND 6-CATEGORY LISTS; PERCENTAGES OF MENTION OF NAMES ON THESE LISTS; AND NUMBER OF *S*s RECEIVING EACH LIST

Categories	Average % Mention of Names	*N*
2 Categories		
Conqueror, Musician	36	37
Artist, Athlete	36	16
Poet, Scientist	36	13
Artist, Conqueror	34	12
Athlete, Poet	37	11
Artist, Poet	34	11
Musician, Scientist	35	10
Mean	36	110
3 Categories		
Conqueror, Musician, Scientist	36	38
Musician, Poet, Scientist	36	15
Artist, Athlete, Poet	36	12
Mean	36	65
6 Categories		
Artist, Athlete, Conqueror, Musician, Poet, Scientist	38	63

[3] Persons interested in the original tabulations of names and the other derived lists may obtain a microfilm copy of the author's dissertation from the Library of the University of California.

his page of names within 1 min. after the first *S* had turned in his, the intervening interval was approximately the same for all *S*s. This interval was 10 min. after the last *S* had finished classification.

To ensure that a minimum number of items was recognized as belonging in each category by each *S*, data were not used if *S* failed to classify, or erred in classifying, 6 or more of the 24 names on the list. Nineteen per cent of the 6-category group were excluded on this basis, as opposed to 10% of the 3-category group and 9% of the 2-category group. Inclusion of these data would increase the variability of the results and lower the average recall scores of all groups. However, an examination of the data showed that the differences among groups would not be affected.

Recall instructions were as follows: "On the blank sheet of your booklet write as many of the names on the list you classified as you can remember. There were 24 names in all, and there were equal numbers in each category. If you have forgotten any of the categories, you may refer to your instruction sheet." Five minutes were allowed for recall.

Results

Equated lists of names were given different amounts of "similarity" by varying the number of categories into which the fixed number of names was placed. Having assumed that items placed in the same category are "similar," we asserted that the effect of increasing the number of categories should be to reduce the amount of competition among the items. According to hypothesis, then, the 6-category group should recall more names than the 3-category group, and the 3-category group should recall more names than the 2-category group. Furthermore, in most cases the percentage of 6-category *S*s recalling a specific name should be greater than the percentage of 3- and 2-category *S*s recalling the same name. Finally, these results should be obtained not only in the gross averages based on the performances of all 6-, 3-, and 2-category *S*s, but also in the averages based on their performances with respect to Artists alone and with respect to each of the other category types.

TABLE 3

Names and Their Percentages of Mention for the 6-Category List

Category	% Mention	Category	% Mention
Artists		*Musicians*	
Michelangelo	60	Bach	74
Picasso	58	Debussy	38
El Greco	18	Mendelssohn	38
Goya	12	Verdi	20
Mean	37	Mean	43
Athletes		*Poets*	
Ruth	75	Poe	69
Dempsey	31	Keats	32
Hogan	20	Chaucer	26
Kramer	13	Sandburg	17
Mean	35	Mean	36
Conquerors		*Scientists*	
Napoleon	88	Newton	51
Genghis Khan	24	Curie	47
J. P. Jones	33	Galileo	29
Bolivar	10	Darwin	16
Mean	39	Mean	36

Grand Mean = 38

Table 4 presents the gross averages and significance of differences. The difference between the 6- and 2-category groups is significant beyond the .001 level of confidence and the differences between the 6- and 3-category groups and between the 3- and 2-category groups are significant by the one-tailed test beyond the .01 level of confidence. Thus the over-all results are according to hypothesis.

The second part of the demonstra-

TABLE 4

Mean Recall of Names by 2-, 3-, and 6-Category Groups

Group	*N*	Mean	*t*	*p*
2-category	110	11.75	2.61 (2–3)	.01
3-category	65	13.05	2.62 (3–6)	.01
6-category	63	14.42	5.68 (2–6)	.001

tion of over-all differences is given by comparison of the recall of the same names by 2-, 3-, and 6-category *S*s. Of the 24 names common to the 6- and the various 2-category lists, 79% were recalled by a higher proportion of 6-category *S*s. Of the 19 names common to the 6- and 3-category groups, 79% again were recalled by a higher proportion of the former. Of the 48 names common to the 3- and 2-category groups, 63% were recalled by a higher proportion of the 3-category group. In general, the differences were greater for the less prominent names (Bolivar) than for the very prominent names (Napoleon).

Another over-all result in accord with the hypothesis is furnished by the intrusion data. An intrusion was said to occur when *S*, in attempting to recall the names that he had previously classified, included a name which had not actually been a member of his list. For example, a 6-category *S* who had classified the musicians Bach, Mendelssohn, Debussy, and Verdi might during the recall period write the names of Beethoven, Mendelssohn, and Debussy. Beethoven would be counted as an intrusion. Our hypothesis was that intrusions would increase as the number of categories decreased, because the greater interference among the 12 items of the 2-category group, for example, would allow more opportunity for strong items not on the list to intrude themselves than in the case of the 4 items of the 6-category group.

The results were that the 2-category group made an average of .75 intrusions, the 3-category group an average of .51 intrusions, and the 6-category group an average of .44 intrusions.

The third part of the analysis of differences in recall is presented in Table 5, which shows the means and significance of differences by specific category. Since the number of items in any one category was different for the 2-, 3-, and 6-category groups, recall percentages were obtained from the raw scores. As the estimate of variance which we wished to use in the *t* test assumes the same variance in both universes sampled, the arc sin transformation was used, whereby percentages are transformed into corresponding angles (11).

TABLE 5

RECALL BY SPECIFIC CATEGORY FOR 2-, 3-, AND 6-CATEGORY GROUPS IN AVERAGE TRANSFORMED PERCENTAGES

Category	No. of Categories	Mean	*df**	*t*†	Comparison
Artist	2	45.39	49	—	2–3
	3	43.70	73	1.71	3–6
	6	53.57	100	2.38	2–6
Athlete	2	43.32	37	—	2–3
	3	46.21	73	1.97	3–6
	6	54.76	88	3.02	2–6
Conqueror	2	43.69	85	2.43	2–3
	3	50.57	99	1.33	3–6
	6	54.86	110	4.54	2–6
Musician	2	42.04	98	3.15	2–3
	3	49.94	114	—	3–6
	6	50.00	108	2.76	2–6
Poet	2	47.14	60	—	2–3
	3	46.28	88	—	3–6
	6	48.81	96	—	2–6
Scientist	2	44.82	74	—	2–3
	3	44.63	114	2.61	3–6
	6	51.79	84	1.90	2–6

* The *df* figure is two less than the number of *S*s of the two groups being compared who categorized names of the specific category under consideration. In the first instance, 16 2-category *S*s categorized Artists and Athletes, 12 categorized Artists and Conquerors, and 11 Artists and Poets, a total of 39. There were 12 3-category *S*s who categorized Artists. (See Table 2.) Thus *df* equal 49.

† By the one-tailed *t* test for 60 or more *df*, a *t* of 1.671 is significant at the 5% level of confidence, and a *t* of 2.390 is significant at the 1% level of confidence.

Table 5 shows that of 18 possible comparisons, 15 are to the advantage of the group with the larger number of categories. There are 10 differences significant at not less than the

5% level of confidence, and all of these are in the predicted direction.

However, there remain the comparisons, particularly several between the 2- and 3-category groups, for which the advantage of a greater number of categories was not demonstrated. These results may be attributed simply to "chance," though several lines of evidence suggest that they are due to uncontrolled differences in the amount of interference which *S*s encountered from the store of Poets, Artists, etc. which were not on the list.

For example, a correlation of .714 between relative size of "parent category" (estimated from the average number of names which the preliminary group of *S*s could think of in 5 min.) and the difference between the recall of the 2- and 3-category groups shows that the 2-category group was inferior when the parent category was large but overtook the 3-category group when the parent category was small.

Our explanation of this result is that in the case of small parent categories the 2-category group was released from most of the proactive inhibition which burdened the recall task of the 3- and 6-category groups. That is, the 12 Artists on the 2-category list comprised most of the Artists known to the *S*s, so that they did not have as much problem in discriminating correct items from incorrect items as did the 3- and 6-category *S*s with 8 and 4 Artists, respectively, on their lists. This explanation assumes amount of interference to be a negatively accelerated function of number of competing items, an assumption that is consistent with the fact that most of the preceding results show more difference between the 6- and 3-category groups than between the 3- and 2-category groups. (Other findings about the relation of number of competing items to amount of interference are conflicting and not strictly relevant to this experiment.)

Reduction of proactive inhibition would also go a long way toward accounting for the results of an experimental variation in which a 1-category group of 13 *S*s was found to recall, not fewer Conquerors or Poets than the 2-category group, but just as many as the 6-category group. The 24 items on this list approximated being all the items of the category that *S* knew, so that interference from incorrect items was greatly reduced. An analysis of the role of proactive inhibition in these experiments is presented more fully elsewhere (7).

Another experimental variation was undertaken to show that "latent categorizing," which we believe occurs almost continually in everyday life, yields the same results as the explicit categorizing that was required of *S*s in the main experiment. Two groups were given the same lists that the 2- and 6-category groups received but were asked only to check whether or not they were familiar with each name. The categories were not mentioned in the instructions.

The results were quite similar to those obtained with explicit categorizing. Eleven 2-category *S*s recalled an average of 10.82 names and 28 6-category *S*s recalled an average of 14.12 names. This difference is significant at the .001 level of confidence. The 2-category list consisted of Conquerors and Musicians, for which previous results had yielded an average of 11.38 for 37 *S*s. The 6-category list, as used in the main experiment, yielded a mean of 14.42 for 63 *S*s.

Discussion

The results of this experiment have limited generality. The number of categories was varied within a small range. Items were distributed equally within categories. The categories were distinct and there was no problem for *S* in remembering the categories. Preliminary experimentation suggests that when these last two conditions do not apply, increasing the number of categories may even decrease the amount of information that can be recalled. Assuming the conditions of the experiment described here, however, we conclude that increasing the number of categories will generally increase the amount that can be recalled.

Understanding why this effect should

"generally" occur draws upon, and serves as a test of, our knowledge of recall processes. Is what has been done here essentially a demonstration of the von Restorff phenomenon (13)? If the von Restorff phenomenon is to be considered as involving figure-ground effects in memory, then this experiment does not seem to fit. It would be difficult to say which was figure and which ground in our "zebra"-like pattern, since the "stripes" were kept equal in size and varied together.

We do interpret this experiment as having shown that different degrees of interference can be created by varying the number of items on a list that are similar in only those abstract attributes that make them members of the same category. However, that this interference comes only from competing tendencies to respond among the experimental items, as the original "crude" analysis stated, was not proven. Since a sample of a population of familiar names served as materials, some of the interference certainly came from proactive inhibition. That is, the larger the number of items of the same category on the experimental list, the more did "recency" discriminanda, and other discriminanda enabling the *S* to distinguish experimental items from other "parent-category" items, generalize to items not on the list, so that the decrement in recall and the large number of intrusions resulted in part, at least, from greater difficulty in discriminating correct items.

Summary

The problem investigated was how the number of categories into which a body of items is classified affects the recall of the items.

The *S*s classified names of famous people into 2, 3, or 6 categories. The different lists of names used were equated on the basis of controlled association data. On recall tests 10 min. after the classification of the names, *S*s who had been given 6 categories recalled significantly more names on the average than those who had been given 3 categories, who in turn recalled significantly more names on the average than those who had used 2 categories. Intrusions varied inversely with number of categories.

The results are interpreted in terms of interference among the items of the experimental list and between those on the list and other familiar items in the *S*'s "parent-categories."

References

1. Bartlett, F. C. *Remembering: a study in experimental and social psychology.* New York: Macmillan, 1932.
2. Bousfield, W. A., & Sedgewick, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.
3. Bruner, J. S., & Austin, G. A. *A theory of cognition.* (Mimeographed.)
4. Carmichael, L., Hogan, H. P., & Walter, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.
5. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
6. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
7. Mathews, R. Recall of items as a function of the number of categories into which they have been classified. Unpublished doctor's dissertation, Univ. of California, 1952.
8. Postman, L. J. Toward a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 242–272.
9. Saul, E. V., & Osgood, C. E. Perceptual organization of materials as a factor influencing ease of learning and degree of retention. *J. exp. Psychol.*, 1950, **40**, 372–379.
10. Siegel, P. S. Structure effects within a memory series. *J. exp. Psychol.*, 1943, **33**, 311–316.
11. Snedecor, G. W. *Statistical methods.* Ames, Ia.: Iowa State Coll. Press, 1946.
12. Tolman, E. C., & Brunswik, E. The organism and the causal texture of the environment. *Psychol. Rev.*, 1935, **42**, 43–77.
13. Von Restorff, H. Ueber die Wirkung von Bereichsbildung im Spurenfeld. *Psychol. Forsch.*, 1933, **18**, 299–342.

(Received July 6, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

PERFORMANCE IN INSTRUMENTAL LEARNING AS A JOINT FUNCTION OF DELAY OF REINFORCEMENT AND TIME OF DEPRIVATION[1]

CHARLES K. RAMOND[2]

State University of Iowa

Initial investigations of behavior often attempt to isolate the effects of single experimental variables. After two or more such influential factors have been identified, it becomes of interest to determine their joint effect, i.e., the manner in which such factors interact or combine to produce response strength.

Two variables which have been observed to affect response speed in simple instrumental learning situations are delay of reinforcement (T_G) (**7, 8**) and time of food deprivation (T_D) (**1, 5, 6, 9**). Hull (**2**) has assumed that T_G and T_D interact multiplicatively, whereas Spence (**10**) recently suggested, on the basis of certain initial evidence (**6**), that the relationship between T_G and T_D is additive.

The present study is a further attempt to determine the manner in which T_G and T_D interact. Specifically, it has investigated a particular response measure, speed of running, as a function of two periods of delay of reinforcement (1 and 5 sec.) and two periods of food deprivation or levels of drive (22 and 4 hr.).

METHOD

Subjects.—The *S*s were 40 male hooded rats from the colony of the Iowa Psychological Laboratory. All were experimentally naive and between 95 and 134 days old at the start of the experiment. Half of the animals were always run 22 hr. after feeding and the other half 4 hr. after feeding. Similarly, half of each drive group received reinforcement 1 sec. after responding, while the other half were reinforced after 5 sec. Thus, the four experimental groups may be identified in terms of their respective degrees of each variable as follows: 22 hr.—1 sec., 22 hr.—5 sec., 4 hr.—1 sec., and 4 hr.—5 sec.

In addition, *S*s were assigned to ten levels according to their ages and litters, such that one *S* from each of the ten levels was in each experimental group. Other potential sources of variance, such as living cage, feeding cage, and order of running, were confounded with this levels factor.

Apparatus.—The apparatus used was a modification of a particular type of Skinner box which has been described elsewhere (**6, 7**). It consisted of a small, triangular-shaped response chamber with flat black, plywood walls and a brass floor. A starting alley led into the front apex through an outer metal door and an inner door of glass. A brass bar could be presented ¾ in. through a hole in the opposite wall. Raising the glass door started a Standard Electric Timer. By touching the bar, *S* completed a low-ampere contact relay circuit (**3**) between the brass floor of the box and the bar. Closure of this circuit stopped the timer, retracted the bar, and fired a Hunter Decade Interval Timer (**4**). One or 5 sec. later, a .08-gm. pellet of food dropped into a food cup situated just below where the bar had been. A 1000-cycle tone sounded concurrently with the release of the pellet and was turned off manually after *S* obtained the food. The apparatus was situated in a dark, sound-resistant room. A shielded 7-w. bulb was located above the food cup; the total illumination level in the response chamber was such that the food cup could not be seen by *E* (nor, presumably, by *S*) unless the 7-w. light above it was on.

Procedure.—All *S*s were placed on a 22-hr. feeding schedule for seven days prior to preliminary training, during which time they were fed 8 gm. of dry Purina Laboratory Chow in individual feeding cages at the same hour every day. The 22-hr. and 4-hr. groups were later fed at different times in order that they could be run at approximately the same hour of the day.

[1] The writer is indebted to Professor Kenneth W. Spence, who suggested the problem and supervised its investigation.

[2] 1st Lt., U. S. Army. Now Acting Military Chief, Human Research Unit No. 3, Office, Chief of Army Field Forces, Fort Benning, Georgia.

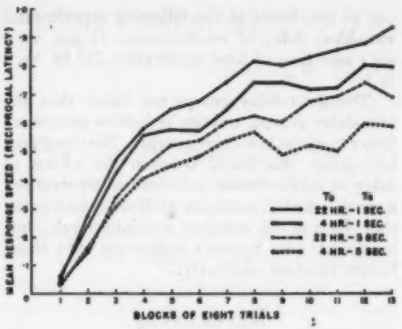

FIG. 1. Running speed as a function of time of deprivation (T_D) and delay of reinforcement (T_G)

Beginning on Day 8, each *S* received 1 hr. of preliminary training daily for three days in order to eliminate irrelevant exploratory behavior during learning, and to establish the secondarily reinforcing properties of the 1000-cycle tone. The bar was never present during this preliminary training. Each *S* was placed in the response chamber with the bar light on and a pellet of mash in the food cup. Thirty seconds after *S* obtained this food and was facing the food cup, another pellet was delivered, accompanied by the sound of the 1000-cycle tone. The tone was turned off after *S* appeared to have eaten the pellet. In this manner 4 more pellets were given at approximately 30-sec. intervals, whereupon *S* was removed from the apparatus until the following day. Preliminary training thus consisted of the receipt of 15 pellets of food, 5 per day.

Training proper began the day after preliminary training, and continued every day thereafter for 16 days. On each trial, *S* was placed in the starting alley and the metal door raised as soon as *S* faced the response chamber; 3 sec. later the glass door was raised, permitting entrance into the response chamber. A total of 104 trials was administered at the rate of four per day for the first six days, and eight per day thereafter. The average intertrial interval was 5 min. Latencies were recorded to the nearest .01 sec. on all trials.

RESULTS AND DISCUSSION

Performance curves for the four groups are shown in Fig. 1. Each point represents the mean of the speeds (reciprocal latencies) of ten *S*s over eight trials. A mean "asymptotic" response speed was calculated for each *S* by averaging his speeds on the last 32 trials. The group means of these values are shown in Table 1, and a summary table of an analysis of variance of this 2 × 2 factorial design is shown in Table 2.

TABLE 1

GROUP MEAN RESPONSE SPEEDS, LAST 32 TRIALS

Time of Deprivation	Delay of Reinforcement		Row Means
	1 sec.	5 sec.	
22 hr.	.885	.709	.797
4 hr.	.772	.559	.666
Column means	.829	.634	

The short-delay groups ran faster ($p < .02$, single-tailed test) than did the long-delay groups, and the high-drive groups ran faster ($p < .08$, single-tailed test) than did the low-drive groups. More importantly, no significant interaction was found between the effects of drive and delay of reinforcement. This latter result may be interpreted as questioning Hull's assumption that T_G and T_D interact multiplicatively, and as supporting Spence's proposal of an additive relationship between T_G and T_D. It should be borne in mind, however, that this finding of no interaction between T_G and T_D may not be generalized with any confidence beyond the particular values of each variable that were employed, i.e., 22-hr. and

TABLE 2

ANALYSIS OF VARIANCE OF MEAN RESPONSE SPEEDS, LAST 32 TRIALS

Source	*df*	Mean Square	*F*	Approx. *p*
Deprivation (T_D)	1	174,240	3.30	.08
Delay (T_G)	1	378,691	7.17	.02
Drive × delay	1	3,423	<1	
Within cells	36	52,844		

4-hr. food deprivation, and 1-sec. and 5-sec. delay of reinforcement. That the additive relationship probably does *not* hold throughout the entire ranges of these factors is clearly indicated by consideration of the joint effects of extreme values of the variables. For example, performance does not improve under conditions of infinite delay of reinforcement no matter how great the time of deprivation employed. Thus, it may be tentatively concluded that the manner in which T_G and T_D combine is more complex than either simple addition or multiplication, but that it may resemble each of these forms of combination, depending upon the particular values of T_G and T_D.

A three-dimensional analysis of these same data indicated that the variance due to the above-described levels factor (a confounding of age, litter, order of running, etc.) was too small to reduce the error term appreciably; hence the data were analyzed simply as above. This similar performance of the different levels may have been the result of a cancellation of opposite effects of the different components which together comprised the levels factor, or it may simply reflect the lack of influence of each component taken singly. In any case, the results suggest the inefficiency of constituting animal *S*s into the type of levels used here.

Summary

Forty male hooded rats were administered 104 reinforced trials in a simple instrumental learning situation. Four groups each received one of two levels of the following experimental variables: delay of reinforcement (1 sec. vs. 5 sec.) and time of food deprivation (22 hr. vs. 4 hr.).

The short-delay groups ran faster than the long-delay groups, and the high-drive groups ran faster than the low-drive groups. No significant interaction was found between the effects of delay of reinforcement and time of deprivation, a result which is contrary to Hull's assumption that T_G and T_D combine multiplicatively, but consistent with Spence's suggestion that these factors combine additively.

References

1. Horenstein, B. R. Performance of conditioned responses as a function of strength of hunger drive. *J. comp. physiol. Psychol.*, 1951, **44**, 210–224.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
3. Hunter, T. A. An electronic contact relay. *Amer. J. Psychol.*, 1952, **65**, 458–459.
4. Hunter, T. A., & Brown, J. S. A decade-type electronic interval timer. *Amer. J. Psychol.*, 1949, **62**, 570–575.
5. Kimble, G. A. Behavior strength as a function of the intensity of the hunger drive. *J. exp. Psychol.*, 1951, **41**, 341–348.
6. Loess, H. B. The effect of variation of motivational level and changes of motivational level on performance in learning. Unpublished Ph.D. dissertation, State Univ. of Iowa, 1952.
7. Logan, F. A. The role of delay of reinforcement in determining reaction potential. *J. exp. Psychol.*, 1952, **43**, 393–399.
8. Perin, C. T. A quantitative investigation of the delay-of-reinforcement gradient. *J. exp. Psychol.*, 1943, **32**, 37–51.
9. Ramond, C. K. Performance in selective learning as a function of hunger. Unpublished Ph.D. dissertation, State Univ. of Iowa, 1953.
10. Spence, K. W. Current interpretations of learning data and some recent developments in S–R theory. In *Symposium on learning theory and personality*, Univ. of Kentucky Press, in press.

(Received June 26, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

OVERTRAINING AND TRANSPOSITION IN CHILDREN

HAROLD W. STEVENSON AND IRA ISCOE

The University of Texas

The problem investigated here has been discussed by Spence (**9**) and has received the attention of Jackson and collaborators in several studies (**4, 5, 6**). On the basis of their studies with young children trained in the usual two-stimulus transposition situation, Jackson *et al.* (**6**) propose that with increasing amounts of overtraining unstable responses appear, followed by stable relative responses, and finally by stable absolute responses in which *S*s choose the training stimulus rather than the new stimulus. The increase in frequency of relative responses after overtraining has also been reported by Hunter (**3**), and the effects of overtraining have been offered by Jones and Dunn (**7**) as a possible explanation of the high relationship that they found between speed of learning and frequency of relative choice. The present study attempts to determine the effect of overtraining on transposition in children who are older than those used in the above studies and who have been trained in a three-stimulus, rather than a two-stimulus, transposition situation.

METHOD

Subjects.—The *S*s were 140 students in grades four through seven in two parochial schools in Austin, Texas, and ranged in age from 9.2 to 14.2 yr.[1]

Materials.—The apparatus consisted of a gray board containing three holes 1¼ in. apart into which gray boxes of the same shade could be fitted. Five boxes were used: three for the training trials and two additional ones for the transposition trials. The dimensions of the boxes were 2¼ × 2 × 1 in. In the center of the lid of each box was mounted a yellow square, the stimulus to be discriminated. The sides of the stimulus squares on Boxes 1 through 5 were ¾, ½, ¼, 1, and 1½ in.

Procedure.—On the training trials Boxes 2, 3, and 4 were used for all *S*s. The position of the boxes on the board was varied randomly from trial to trial, with the same random sequence being used for each *S*. The *S* was to find the box in which an eraser was hidden. For half of the *S*s it was always to be found in Box 4, and for half in Box 2. These may be designated the Large and Small groups.

Two degrees of correct response were required: half of the *S*s were required to learn to the criterion of four consecutive correct responses, while the other half had to make ten consecutive correct responses. We may designate these the 4-trial and 10-trial groups. As a result of this procedure four groups of 35 *S*s each were formed: Large-4; Large-10; Small-4; and Small-10.

When each *S* entered the testing room he was seated across the desk from *E*. At *S*'s right was a cardboard screen and in front of him was the board with the boxes. The *S* was told: "This is a game we've been playing with boxes. See, each box opens like this. Now see this [eraser]? I'm going to hide this in one of the boxes and I want to see if you can find it." The *E* removed the board from *S*'s sight behind the screen and placed the eraser in one of the boxes. The board was then placed in front of *S* and he was told, "open the box that you think it is in." After each choice the board was placed behind the screen and the boxes rearranged. The trials were continued for *S* until he met the criterion to which he had been randomly assigned.

The test trials followed immediately with no additional instructions or remarks. For the Large groups, Box 2 was removed and Box 5 introduced; and for the Small groups, Box 4 was removed and Box 1 introduced. The test trials continued until four consecutive correct responses were made. The eraser was found in the largest or smallest box, depending upon whether *S* was in the Large or Small groups. When *S* met this criterion, a second set of transposition tests was given. For the Large groups Boxes 4 and 5 were removed and Boxes 1 and 2 introduced; for the Small groups Boxes 1 and 2 were removed and Boxes 4 and 5 introduced. Again, the eraser was

[1] We wish to acknowledge the kind cooperation of Sr. M. Bernarda of St. Mary's School and Sr. M. Eugenia of St. Austin's School in this study.

TABLE 1

TRIALS REQUIRED TO ATTAIN TRAINING CRITERION

Group	N	Age		Trials to 4 Correct Trials		Trials to 10 Correct Trials	
		Mean	SD	Mean	SD	Mean	SD
Large-4	35	11.2	.9	16.7	16.2		
Large-10	35	11.4	.9	21.1	23.7	28.9	27.8
Small-4	35	11.3	1.4	23.6	24.7		
Small-10	35	11.2	1.0	17.5	18.6	26.5	29.3

found in the largest or smallest boxes in this new sequence.

At the conclusion of the experiment *Ss* were asked, "How did you know it was in that one?" and were urged to verbalize their basis of response.

RESULTS

Training trials.—The mean number of trials required by all *Ss* to reach the 4-trial criterion and for the overtrained *Ss* to reach the 10-trial criterion may be seen in Table 1. (Here, and throughout the discussion, the number of trials is *exclusive* of the criterion trials.) The mean age for each group is also shown, indicating that as far as age is concerned, the four groups were quite homogeneous.

The distributions of learning trials were all markedly skewed. Some *Ss* in each group solved the problem immediately by choosing the correct box on the first choice and continuing to choose it correctly to criterion, while others took up to 111 trials to reach the criterion. The great differences in ease of acquisition of a 4-trial criterion for all *Ss* are not, at least within the rather restricted age range used, related significantly to chronological age ($r = -.11$). Mental age ratings were not available.

In order to determine the comparability of the 4- and 10-trial groups in their ability to acquire the size discrimination, we may combine the *Ss* according to learning criterion (Large-10 with Small-10 and Large-4 with Small-4) and test whether the mean number of trials required to attain a criterion of four consecutive correct responses differed significantly. The combined mean for the two 10-trial groups was 19.3, and 20.6 for the 4-trial groups, with corresponding standard deviations of 21.4 and 20.7. Applying Festinger's technique (2) for evaluating the differences between means derived from skewed distributions, an F of 1.07 results ($df = 113, 139$), which is not significant ($p > .05$). We may assume, therefore, that the groups were comparable in their ability to learn such a discrimination.

As may be seen in Table 1, the effect of the more stringent (10-trial) criterion was to increase the number of training trials for *Ss* in these groups, i.e., to result in overtraining for these *Ss*. Comparing the mean number of trials required by the 10-trial and the 4-trial groups to reach their respective criteria (again excluding the criterial trials) an F of 1.34 results ($df = 132, 140$), which is significant at between the .05 and .01 levels.

The mean number of times that each box was chosen up to the criterion trials may be seen in Table 2. It is clear that some overtraining on the correct choice had occurred before the ten criterion trials commenced in the 10-trial groups, for the difference between the means of the 4- and 10-trial groups is notably greater for the correct stimuli than it is for any of the other stimuli. Excluding the criterion trials, the correct stimulus was chosen

on the average 4.2 (11.8 vs. 7.6) more times in the 10-trial groups than in the 4-trial groups. A comparison of the difference between these means is significant at beyond the .01 level ($F = 1.55$; $df = 195, 88$).

Test trials.—If transposition is tested by determining the number of *Ss* who responded relatively on the first test trial, it is found that 83 of the 140 *Ss*, or 60%, showed transposition. Of the remainder, 38 *Ss* chose the stimulus on which they had been trained, and 19 chose the third stimulus. The breakdown of these data is found in Table 2. The percentage of *Ss* showing transposition is notably lower than that usually found in two-stimulus transposition for younger *Ss* similarly tested with stimuli one step removed from the training stimulus. Unfortunately, additional data for *Ss* of this age level or for three-stimulus transposition are not available by which this relatively low frequency might be explained.

Of the 83 *Ss* who showed transposition, 38 were in the 4-trial and 45 were in the 10-trial groups. This difference is not significant ($\chi^2 = 2.4$; $.10 < p < .20$), and we may conclude that the degree of overtraining employed here did not significantly increase the frequency of transposition on the first test trial. Further, there was no difference between the Large and Small groups in their frequency of relative choice ($\chi^2 = .92$). In no case throughout the remainder of the experiment was a significant difference found between these groups, and henceforth the Large and Small groups will be combined.

If we apply a more sensitive test of transposition, that of consistency of response on the test trials, the picture changes radically. Since, in the method used in this experiment all *Ss* were given at least four test trials, we may determine the number of *Ss* who made a relative choice consistently on the first four test trials. The number of *Ss* who may be judged to show transposition according to this criterion is 72, and of these only 28 were in the 4-trial groups, while 44 were in the 10-trial groups. The chi square that results here is 7.31, which with 1 *df* is significant at beyond the .01 level. Further evidence of this difference is provided by the fact that the mean number of trials required (up to, but not including the four criterial test trials) by the 10-trial groups was 1.27 and by the 4-trial groups was 3.87. Again applying the Festinger test, this difference is significant at beyond the .001 level ($F = 3.04$; $df = 32, 157$).

On the first trial of the second set of transposition trials, 116 of the 140 *Ss* made a relational response, an increase in frequency of 23.4% over the first trial of the first set of transposition trials. In this trial the 10-trial groups

TABLE 2

THE *Ss*' CHOICES ON TRAINING AND TEST TRIALS

Group	Training Trials Stimuli (Means)			Test-1 Stimuli (Frequencies)			Test-2 Stimuli (Frequencies)		
	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large
Large-4	5.5	5.5	5.7*	6	9	20*	6	3	26*
Large-10	7.5	8.9	12.5*	3	7	25*	5	0	30*
Small-4	9.5*	7.5	6.6	17*	12	6	26*	5	4
Small-10	11.1*	7.9	7.5	21*	10	4	34*	0	1

* Correct stimulus.

made a significantly greater number of relational choices than did the 4-trial groups. The chi square for frequency of relational versus nonrelational responses in the two groups is 7.24, which is significant at beyond the .01 level ($df = 1$). Utilizing consistency of relational responses as the measure of transposition, we find that 105 *Ss* met this criterion, with 46 in the 4-trial groups and 59 in the 10-trial groups. In comparing these frequencies a significant chi square is also obtained (6.44; $.01 < p < .02$). The choices made on the first trial of the second set of transposition trials are shown in Table 2.

In summary, the frequency with which *Ss* in the 4- and 10-trial groups chose the relative stimulus did not differ significantly on the first trial of the first set of test trials, but was significantly greater in the 10-trial groups on the first trial of the second set of test trials. On both sets of test trials the 10-trial groups showed a significantly greater frequency of consistent relative choices than did the 4-trial groups.

Verbalization.—By the end of the experiment each *S* had been forced to make the same relative choice on at least four training trials and on two sets of test trials, yet only 111 of the 140 *Ss* were able to verbalize their basis of response. Those who did verbalize gave answers varying from terse explanations that it was always the biggest (or smallest) square to detailed descriptions of the procedure for the entire experiment. Those who did not verbalize appropriately said they did not know or made such undefinitive statements as "It's in that box." The nonverbalizers proved to be somewhat younger than the verbalizers, being on the average 9 mo. younger. Seventeen of the nonverbalizers were in the 4-trial groups and 12 in the 10-trial groups.

TABLE 3

CHOICES MADE BY VERBAL AND NONVERBAL *Ss* ON FIRST TRANSPOSITION TRIALS

Type of Response	Class of Subjects	
	Verbal	Nonverbal
Relational	72	11
Nonrelational	39	18
Consistent relational	66	6
Nonconsistent relational	45	23

The nonverbalizers did significantly poorer in transposition than did the verbalizers. Testing the frequency with which the verbalizers and nonverbalizers made relational responses on the first test trial of the first transposition series as seen in Table 3, a chi square of 6.91 results, which with 1 *df* is significant at beyond the .01 level. Testing the difference in number of *Ss* making consistent correct responses on the first test series, a chi square of 14.09 is obtained, which is significant at well beyond the .001 level. The ability to verbalize, then, was positively correlated with the tendency to make relational choices on the test trials.

DISCUSSION

The results of this experiment only partly support the generalizations concerning overtraining that were made by Jackson *et al.* (6), for although overtraining affected the consistency with which relative responses were made, it did not have a significant effect in increasing the initial frequency of relative responses until a second set of test trials was given.

In interpreting this experiment, then, it is proposed that the responses were a result of a conceptual process rather than of a strong tendency towards perceiving situations relationally or of an interaction of habit strengths. This interpretation is supported by the high relationship that was found between ability to verbalize at the end of the experiment and the incidence of transposition. Sim-

ilarly, both Kuenne (8) and Alberts and Ehrenfreund (1) found that those *S*s who were able to abstract from their experience to the degree that they could verbalize appropriately concerning the basis of response showed the greatest frequency of transposition.

We may hypothesize that two concepts had to be developed, a concept of size and of frequency. That is, first the *S*s had to develop some concept about the particular type of stimulus that would be correct, and second a concept about the frequency with which this stimulus would be correct on any series of trials. The primary effect of overtraining apparently was that of establishing a clearer concept of consistency than of relative size. Once the overtrained *S*s found a stimulus to be correct, irrespective of the basis of their choice, they continued to choose the correct stimulus to a greater degree than did the *S*s in the 4-trial groups. The effect of overtraining on initial choice of the test stimuli was not apparent until the second set of transposition trials, when the more complete mastery of the initial discrimination did enable these *S*s to learn more readily the principle of response common to the training and the first set of test trials.

Overtraining, then, may be effective in increasing the frequency of transposition, but in the present experiment its effects seem to have resulted from the expanded opportunity of *S*s to discover that similar principles applied in two series of stimuli with similar relational characteristics. It may be predicted, therefore, that overtraining with the same set of stimuli would have less effect on transposition than if some of the overtraining were given with a different set of stimuli embodying the same principle.

Summary

A group of 70 *S*s was trained to choose the largest of three stimuli and a second group of 70 *S*s was trained to choose the smallest. Half of each group in turn were required to reach a criterion of either four or ten consecutive correct responses. Two sets of transposition trials were made with stimuli removed one step in either direction from the training stimuli. The following results were obtained:

1. The amount of transposition was less than that found in previous studies with younger children trained on two stimuli.
2. No significant differences in transposition behavior were associated with the size of the training stimulus.
3. No significant differences in number of relative choices on the first trial of the first transposition series were found between the groups differing in degree of original training, but significant differences favoring the overtrained groups did arise on the first trial of the second set of test trials.
4. In both sets of test trials the overtrained *S*s chose the correct stimulus in a consistent sequence to a greater degree than the less well-trained *S*s did.
5. Appropriate verbalization was found to be positively related to the ability to show transposition.

References

1. Alberts, E., & Ehrenfreund, D. Transposition in children as a function of age. *J. exp. Psychol.*, 1951, **41**, 30–38.
2. Festinger, L. A statistical test for means of samples from skew populations. *Psychometrika*, 1943, **8**, 205–210.
3. Hunter, I. M. L. An experimental investigation of the absolute and relative theories of transposition behavior in children. *Brit. J. Psychol.*, 1952, **42**, 113–128.
4. Jackson, T. A., & Dominguez, K. Studies in the transposition of learning by children: II. Relative vs. absolute choice with with multi-dimensional stimuli. *J. exp. Psychol.*, 1939, **24**, 630–639.
5. Jackson, T. A., & Jerome, E. Studies in the transposition of learning in children: IV. A preliminary study of patternedness in discrimination learning. *J. exp. Psychol.*, 1940, **26**, 432–439.
6. Jackson, T. A., Stonex, E., Lane, E., & Dominguez, K. Studies in the transposition of learning in children: I. Relative vs. absolute response as a function of the amount of training. *J. exp. Psychol.*, 1938, **23**, 578–600.
7. Jones, H. E., & Dunn, D. The configural factor in children's learning. *J. genet. Psychol.*, 1932, **41**, 3–15.
8. Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
9. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received June 29, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

EFFECTS OF NONREINFORCED TRIALS ON DISCRIMINATION LEARNING IN PRESCHOOL CHILDREN

GORDON N. CANTOR AND CHARLES C. SPIKER[1]

Child Welfare Research Station, State University of Iowa

Various approaches to discrimination learning phenomena, such as that represented by Hullian theory (2) and the Spence discrimination theory (3), maintain that learning occurs in a discrimination learning situation owing to the nonrewarding of incorrect responses, as well as to the rewarding of correct responses, provided that an initial tendency exists to approach the incorrect stimulus. In an investigation of discrimination learning, Denny and Dunham (1) found that a group of rats given four nonreinforced and two reinforced trials per day learned a discrimination task in a simple T maze significantly better than did a group given one nonreinforced and two reinforced trials per day. The present study is designed to investigate the effect of manipulation of the frequency of nonreinforcement on the performance of preschool children on a forced-choice discrimination learning task.

METHOD

Subjects.—The *S*s were 48 children from the State University of Iowa preschools. There were 13 three-year-olds, 18 four-year-olds, and 17 five-year-olds; 26 were males, and 22 were females.

Procedure.—Figure 1 shows the main part of the apparatus. Tracks were arranged down which *S* could run a wooden railroad car from a standard toy set. Two such cars were used; these were identical in appearance, except that one was painted red, the other green. Each child was run individually. The *S* stood facing the side of the tracks with the high ends at his left. The experiment began with a demonstration series in which *E* rolled each of the cars down one of the tracks. Upon contact of a car with the dummy switch at the end of a track, *E* activated the marble dispenser by means of a hidden foot switch, causing the marble dispenser to roll a ⅝-in. steel bearing into a receptacle accessible to *S*'s right hand. The *S* was led to believe that the car's striking the dummy switch was what caused the marble dispenser to operate, and was told to place the marbles obtained during the demonstration series in a container which lay on the table behind the tracks. He was also informed that the object of the "game" was to fill this container with marbles.

The *E* then began presentation of the discrimination learning task to *S*. For each *S*, free choice trials were alternated with blocks of forced trials. The task began in every case with a free choice trial in which both cars were presented on the delivery mechanism to *S*, who was instructed to "take just one." The *S* chose one of the cars and rolled it down either of the tracks. If this choice was to be reinforced, *E* caused the dispenser to roll a marble to *S;* if the choice was not to be reinforced, no marble was delivered. Next, *S* was given his first block of forced trials; these consisted in the presentation to *S* of only one stimulus at a time, one color being consistently reinforced for the given *S*, while the other was consistently nonreinforced. The alternation of single free choice trials with blocks of forced trials was continued in exactly this manner until *S* had been given nine free choice trials and eight blocks of forced trials.

The manipulated variable of basic interest involved the number and kind of forced trials given in each of the eight blocks. For half of the *S*s, two rewarded and two nonrewarded forced trials were given in each block, while for the other half, two rewarded but only one nonrewarded forced trial was given per block. It was predicted that the former group would surpass the latter in number of correct choices made on the free choice trials.

For half of each of the above-described groups, all free choice trials were reinforced, regardless of correctness or incorrectness of choice, while for the other half, all free choices were nonreinforced.

[1] The writers are indebted to Mr. Robert Seibel for technical assistance.

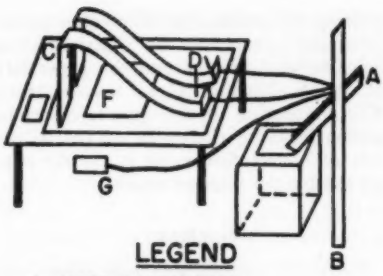

FIG. 1. Schematic drawing of apparatus

RESULTS AND DISCUSSION

Taking into account the two variables of amount of nonreinforcement during forced trials and presence or absence of reinforcement on free choice trials, four treatment groups are seen to be involved, as follows: (I) forced trials reinforced and nonreinforced equally, free choice trials reinforced; (II) forced trials reinforced and nonreinforced equally, free choice trials nonreinforced; (III) forced trials reinforced and nonreinforced unequally, free choice trials reinforced; and (IV) forced trials reinforced and nonreinforced unequally, free choice trials nonreinforced. The measure of interest in the analysis is the number of correct choices made out of the last eight of the nine free choice trials. The means and sigmas of the four treatment groups are shown in Table 1.

TABLE 1

FREQUENCY OF CORRECT CHOICES ON FREE CHOICE TRIALS

Group	Subjects			Mean Age (Months)	Correct Choices	
	Total	Males	Females		Mean	*SD*
I	12	9	3	50.42	5.83	1.57
II	12	4	8	49.50	6.08	1.71
III	12	7	5	53.42	4.17	1.57
IV	12	6	6	50.33	5.08	1.38

TABLE 2

ANALYSIS OF VARIANCE OF FREE CHOICE DATA

Source	*df*	Mean Square	*F*
Reinforced vs. nonreinforced free choices	1	4.09	1.54
2 vs. 1 nonreinforced trials per block	1	21.34	8.02*
Interaction	1	1.32	<1.00
Within cells	44	2.66	
Total	47		

* $p = .01$.

The basic analysis of the data involved a 2×2 factorial design which compared the performance of *Ss* with two nonreinforcements per block of forced trials with that of *Ss* with only one nonreinforcement per block, and compared the performance of *Ss* having all free choices reinforced with that of *Ss* having all free choices nonreinforced. Table 2 summarizes the results of the analysis of variance.

The interaction is nonsignificant, as is the main effect for reinforcement vs. nonreinforcement of the free choice trials. For the main effect of equal vs. unequal reinforcement and nonreinforcement in blocks of forced trials, however, the obtained *F* of 8.02, with 1 and 44 *df*, is significant at the .01 level. That is, the data indicate that negative instances forced upon *S* during training of this nature aid in the learning process. The combined mean for *Ss* reinforced and nonreinforced equally in forced trials (Groups I and II) is 5.96, while the mean for *Ss* reinforced and nonreinforced unequally (Groups III and IV) is 4.62. It apparently made no difference in the learning whether *E* arbitrarily rein-

forced or nonreinforced all free choices. The combined mean for *S*s reinforced on all free choice trials (Groups I and III) is 5.00, while the mean for *S*s nonreinforced on free choice trials is 5.58.

A chi-square analysis of the data for the first trial revealed no significant position-color preference in the *S*s. A recording of any remarks made by the *S*s concerning a color or colors involved in the experiment reveals that, of the 48 *S*s, 28 did make some mention of color, whereas 20 did not. A comparison of these two groups in terms of number of correct choices on the last eight free choice trials resulted in a t value of 2.02 which, for 46 df, is significant at the .05 level. Inspection of the data, however, indicates a marked tendency for those *S*s to mention color who were given as much nonreinforcement as reinforcement in the training trials. This fact detracts from the importance of this comparison of color verbalizers and color nonverbalizers, since nothing can be said concerning this factor independent of the main treatment effects.

The simple analysis of variance comparing the three age groups (3-, 4-, and 5-year-olds) resulted in an F of less than 1, indicating no difference in performance on the part of these age groups.

The data agree with the results obtained by Denny and Dunham (1) concerning the enhancing effect of negative "training" instances on discrimination learning and thus extend the generality of the findings of these investigators. The contention of the Spence discrimination learning theory in regard to the role of nonreinforcement is clearly supported by the present results.

Summary

Forty-eight children, ages three to five years, were run in a forced choice simple discrimination learning situation in which blocks of forced choices were alternated with free choice trials. Half the *S*s received two reinforced and two nonreinforced trials in each block of forced trials, while the other half received two reinforced but only one nonreinforced trial per block. The performance of the former group on the free choice trials was significantly superior to that of the latter group, indicating the importance of nonreinforcement in discrimination learning.

References

1. Denny, M. R., & Dunham, M. D. The effect of differential nonreinforcement of the incorrect response on the learning of the correct response in the simple T maze. *J. exp. Psychol.*, 1951, **41**, 382–389.
2. Hull, C. L. *A behavior system.* New Haven: Yale Univ. Press, 1952.
3. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

(Received July 18, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

RESPONSE CONSISTENCY IN PERCEPTION AND RETENTION

EUGENE S. GOLLIN AND ALAN BARON

Brooklyn College

Recent experimental studies by Witkin (**17, 18**) and Klein (**6**) have demonstrated the existence of wide individual differences in a variety of perceptual situations. It was also found that individuals tended to respond in a consistent fashion in the various perceptual situations. This consistency of *response style*, as well as the consistency of differences between *S*s, has permitted Witkin and Klein to characterize *S*s in terms of perceptual performance. If performance in perceptual situations is characteristic of individuals, that is, reflective of central response tendencies, then it should relate to performance in behavioral situations generally regarded as being nonperceptual. This relationship should be especially evident when the representative test situations have common organizational properties.

Gestalt psychologists have likened the processes involved in retention to those operating in visual perception. They have described retroactive inhibition (RI) situations in terms of figure-ground organization and in this way have accounted for the superior retention of dissimilar elements in otherwise homogeneous material. The distinctiveness existing between successively learned materials influences their later retention in the same way that the distinctiveness existing between aspects of the visual field influences their perception (**7**). Other investigations of the RI phenomenon have shown that varying the degree of distinctiveness between successively learned materials (overlearning, dissimilarity, etc.) will facilitate the later reproduction of the originally learned material (**2, 4, 8, 9, 10, 11, 12, 13, 15**).

A perceptual situation which permits manipulation of distinctiveness between aspects of the visual field is the Gottschaldt test. Discovery of the hidden figure in this test involves a discrimination between a simple figure with which *S* has had previous experience and the context in which it is embedded. Embeddedness of the hidden figure can be accomplished in varying degrees in much the same way that similarity between successively learned materials can be manipulated in RI situations.

In both Gottschaldt and RI, *S* is directed to produce material which is obscured by experimental manipulation of stimulus conditions. It has been demonstrated that when these conditions are held constant for all *S*s in the Gottschaldt, wide individual differences in performance emerge (**14, 18**). Individual differences are also forthcoming in RI situations.

In each of these tasks, performance is related to the ease with which the *figural* material is maintained in the face of hindering field conditions. When test conditions are held constant for all *S*s, resulting differences in performance may reasonably be ascribed to individual differences in the facility with which a figure can be maintained against a ground. If *S*s perform consistently in Gottschaldt and RI, evidence will be provided for the pervasiveness of this facility.

METHOD

Subjects.—The *S*s were 25 male students randomly selected from introductory psychology

classes and assigned to a control group ($N = 5$) and an experimental group ($N = 20$).

Materials.—The Gottschaldt test consisted of five sets of figures, one of which was used for practice. Each set contained a simple inspection figure and seven complex figures in which the simple figure was embedded. Thus, in addition to the practice set there were 28 complex figures. The figures in the test were taken from the Gottschaldt originals, from those used in the experimental psychology laboratory at Brooklyn College, and were in part independently constructed.

In the RI situation two lists of nonsense syllables were used. There were ten syllables in each list and they had Glaze association values between 20% and 33% (5).

Procedure and design.—The *S* was presented with the practice Gottschaldt set, and after inspecting the simple figure was requested to outline it in pencil on the seven complex figures. In the test proper, *S* inspected the simple figure for 10 sec. and then attempted to outline it on the successively presented complex figures. The *S*s were timed to the nearest second on each complex figure. Timing was stopped if *S* incorrectly outlined the simple figure and he was informed of his error. The *S* was allowed to inspect the simple figure at any time in the course of the testing. During such inspection, timing was also stopped. At no time were *S*s permitted to see the simple and complex figures together. The *S*s were allowed a maximum of 60 sec. in which to solve each complex figure.

In the RI situation the nonsense syllables were presented in an electrically driven circular memory drum at a constant rate of speed, each syllable being exposed for 2 sec. As each syllable appeared, *S* pronounced it aloud. Mispronunciations were corrected on the first trial. At the end of each trial, *S*s were given a 4 × 6-in. file card and requested to write as many of the syllables as they could in the same order as presented and with the same spelling. The cards were removed as soon as *S* had finished writing. This procedure was repeated after each trial until *S* was able to write the list twice consecutively without error. To compensate for the lack of a practice series, learning was preceded by a demonstration of the apparatus and extensive instructions.

The control group served only in the RI situation. Both experimental and control groups learned the first list of syllables. Experimental *S*s then learned the second list while control *S*s rested. The rest period was presented as a test in which control *S*s were required to select from a book of cartoons the three they liked best. The *S*s in both groups were then tested for recall by having them write as many syllables of the first list as they were able to produce. No more than 1 min. was allowed for this. The *S*s were then required to relearn the first list to a criterion of two perfect trials.

The tests were administered in counterbalanced order, half the experimental *S*s taking the Gottschaldt first and half participating in the RI situation first.

Results

In the RI situation List I (originally learned list) and List II (interpolated list) were of approximately equal difficulty. The mean number of trials required to learn List I was 11.90 as compared with 10.40 for List II ($t = 1.45$, $p < .20$).

The control and experimental groups were adequately matched with regard to original learning. A statistical comparison of the mean number of trials required by each group to reach the criterion yielded a $t = .83$ ($p > .40$).

The interference effect of List II is demonstrated by an RI percentage of 74 obtained by use of the standard formula (1), mean recall for the control group being 8.4 syllables as compared with 2.2 syllables for the experimental group.

In the Gottschaldt test, mean time scores ranged from 5.39 sec. to 29.75 sec. The mean of the mean time scores was 14.59 sec. Although *S*s differed markedly from one another in terms of mean time scores, they remained fairly self-consistent throughout the trials as indicated by a correlation of .81 between the 14 odd and 14 even trials.

In order to test the hypothesized pervasiveness of figure-ground facility, it was necessary to compare performance in the Gottschaldt test with performances in the learning situations. Gottschaldt time scores were converted to speed scores by reciprocal transformation (3) and rank-order correlations were run between Gott-

TABLE 1

RANK-ORDER CORRELATIONS

Measure	Saving Score	Relearning Trials	Recall Score	Original Learning Trials
Gottschaldt speed (1/time)	.58*	−.51**	.45**	−.08
Relearning trials			−.74*	.28
Recall score	.82*			−.08

* $p < .01$.
** $p < .05$.

schaldt speed and number of trials required for relearning, number of trials required for original learning, number of syllables achieved on the first relearning trial (recall score), and savings score. Other relevant rho's were also obtained.

Inspection of Table 1 indicates that the magnitude of the correlations between Gottschaldt speed and relearning trials, recall scores, and saving scores are significantly different from zero. These results show that the test situations are productive of individual differences and that *Ss* who are faster on the Gottschaldt test tend to be less susceptible to the interference in the RI situation. It should be noted that while a relationship significantly different from zero exists between Gottschaldt and recall, no such finding was obtained between either of these and original learning trials. Furthermore, the rho between original learning trials and relearning trials is not significantly different from zero. Gottschaldt, recall scores, and relearning trials seem to test the stability with which previously experienced material is maintained under conditions of interference, while original learning trials seem to test the rate of establishment of material.

DISCUSSION

The results of this study confirm Witkin's findings of a wide array of individual differences in the Gottschaldt. This finding, as Witkin points out, ". . . proves that it is necessary to consider characteristics of the people perceiving the figures as well as characteristics of the figures themselves. The observation that people tend to be self-consistent in the ease or difficulty with which they escape the influence of the complex pattern points to the importance of the personal factors responsible for individual differences" (18, p. 13).

In the retention tests utilized in this study marked differences between individuals were also found, indicating that the number of trials required for relearning as well as recall as measured by number of syllables produced on the first relearning trial depended not only on the nature of the learning tasks but also on personal factors.

A consideration of the structural properties of the tasks involved in this experiment suggested that performance was related to the facility with which figural material could be produced in the face of hindering field conditions.

The fact that *Ss* who are low on relearning trials and high on recall scores tend to locate hidden figures with greater speed suggests that this facility is pervasive. This demonstration of pervasiveness of response style, is, of course, confined to the conditions of the experiment. Moreover, no statement can be made regarding the source of the differences between individuals. It is felt, however, that motivational variables may contribute to figure-ground facility.

The experimental findings suggest that ability to make distinctions between aspects of given fields may represent a useful personological dimension. In this regard, the presence of figure-ground relationships outside the perceptual area has been noted by Woodworth (19). Werner has utilized a similar conception in a study of proactive inhibition (16).

SUMMARY

A retention task (RI situation) and a perception task (modified Gottschaldt test) having common organizational properties were pre-

sented to *Ss* in order to determine whether they would respond consistently in the two test situations. Evidence of consistency of response was forthcoming.

References

1. Andrews, T. G. (Ed.) *Methods of psychology.* New York: Wiley, 1948.
2. Britt, S. H. Retroactive inhibition: a review of the literature. *Psychol. Bull.*, 1935, **32,** 381–440.
3. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28,** 93–115.
5. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–267.
6. Klein, G. S. The personal world through perception. In R. R. Blake & G. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951.
7. Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.
8. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
9. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53,** 173–203.
10. Robinson, E. S. The "similarity" factor in retroaction. *Amer. J. Psychol.*, 1927, **39,** 297–312.
11. Robinson, E. S., & Heron, W. T. Results of variations in the length of memorized material. *J. exp. Psychol.*, 1922, **5,** 428–448.
12. Sand, M. C. The effect of length of list upon retroactive inhibition when degree of learning is controlled. *Arch. Psychol.*, 1939, **33,** No. 238.
13. Skaggs, E. B. Further studies in retroactive inhibition. *Psychol. Monogr.*, 1925, **34,** No. 8 (Whole No. 161).
14. Thurstone, L. L. *A factorial study of perception.* Chicago: Univ. of Chicago Press, 1944.
15. Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59,** No. 3 (Whole No. 273).
16. Werner, H. The effect of boundary strength on interference and retention. *Amer. J. Psychol.*, 1947, **60,** 598–607.
17. Witkin, H. A. The nature and importance of individual differences in perception. *J. Pers.*, 1949, **18,** 145–170.
18. Witkin, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, **19,** 1–15.
19. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received July 6, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

APPARENT SPATIAL ARRANGEMENT AND PERCEIVED BRIGHTNESS

JULIAN E. HOCHBERG
Cornell University

AND

JACOB BECK
Yeshiva College

The problems of brightness constancy (e.g., the constancy of perceived object color under different illumination conditions), and of the perceptual constancies in general, arise from the fact that changed sensory stimuli frequently elicit unchanged responses (and vice versa) which follow more closely the variations of distal stimuli (objects) than of the sensory-surface stimulus distributions. This raises difficulties for any formulation of a one-to-one correspondence between stimulus and experience (confusingly called the "constancy hypothesis," [**5**, p. 86]), which at first sight would seem essential to psychological prediction.

Such findings have been used in attempted ("nativistic") refutation of the constancy hypothesis and its associated stimulus-sensation units of analysis (**5**), and to demonstrate the importance of nonstimulus organizational "forces." Empiricist "inferential" explanations, on the other hand, retain the constancy hypothesis in *sensation*, and ascribe the obtained discrepancies to the effects of past experience in *perception*. Objections to the nativistic position are: (*a*) some evidence suggests that the accuracy of the perceptual constancies depends on past experience (**1**); (*b*) no well-defined analytic units have been presented to supplant the old "sensations," and in their absence precise prediction is difficult despite the considerable heuristic value of the more or less intuitive Gestalt "laws." General objections to the empiricist positions have been: (*a*) it is not possible to distinguish between "sensation" and "perception"; (*b*) there is some awkwardness involved in the doctrine of "unconscious inference" and its derivatives, especially when referred to the lower animals in which the constancies appear (**7**, pp. 605–607); (*c*) any attempt at precise prediction from this viewpoint must await as yet unperformed "ecological surveys" to determine what the past experiences of an organism are likely to have been; (*d*) the constancies also appear to exist without opportunity for past experience (**3**), and while demonstrated effects of past experience on the constancies do not necessarily refute the Gestalt position, evidence of the reverse seriously injures a thoroughgoing empiricist explanation.

An alternative formulation is appealing: responses may occur in one-to-one correspondence not to what we had previously taken to be the stimuli but to their relationship, without regard to central factors, whether of association or organization. In considering this possibility, we do not have to postulate innate knowledge; we need only seek new dimensions for analyzing the physical stimuli which *are* in correspondence with experience (or response). Gestaltists most frequently sought such invariant relationship not in the stimulus distribution, but in the as yet largely unmeasurable psychophysiologically isomorphic cortical processes; however, one may instead direct attention to the reanalysis of the proximal stimulus pattern as do Gibson (**2**), Helson (**4**), and Wallach (**6**).

Thus, Wallach (**6**) suggests that we

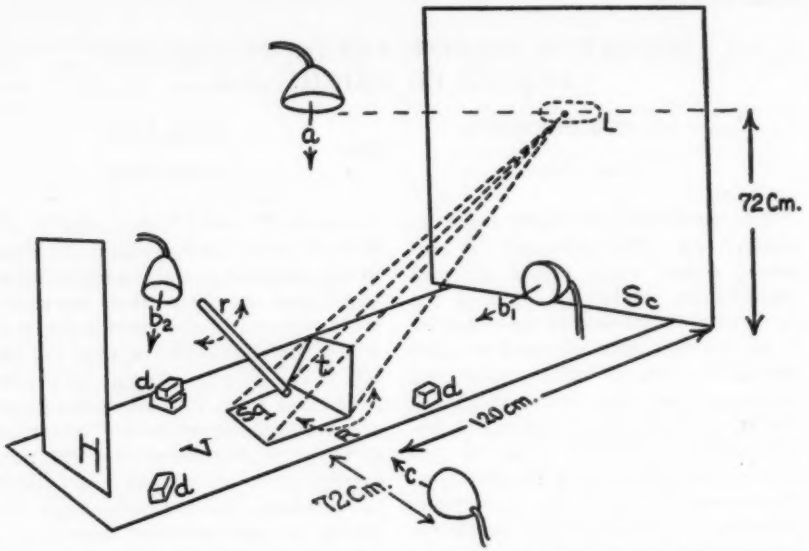

FIG. 1. Apparatus for presenting the same target at different apparent slants and illumination conditions

may understand brightness perception by taking as the stimulus not the intensity of illumination falling on a given retinal region, but the relationship of the intensities of illumination falling on adjacent regions. The relationship approximated a *ratio* of intensities in the situations he studied; i.e., the stimuli in the perception of brightness appeared to be the ratios of illumination intensities on adjacent areas, rather than the illumination intensities themselves. Thus, if *Ss* viewed a variable disk surrounded by a ring of 180 illumination intensity units—degrees of episcotister opening—and were asked to match the variable disk to a disk of 90 units surrounded by a ring of 360 units, they set the variable disk to a mean value of 47 units, only 2 units away from the proportionate 4:1 intensity ratio, which would here be $90 \div 360 \times 180$, or 45 units. If this redefinition of the stimulus will explain all of brightness constancy, we can again effect a one-to-one psychophysical formulation of perceived brightness.

The constancy hypothesis was shaken since the same *absolute* stimulus intensities aroused different brightness responses (and vice versa); can conditions also be found in which the same *distributions* or relationships of stimulus intensity arouse different brightness responses? The object of the present experiments was to determine

TABLE 1

RELATIVE APPARENT BRIGHTNESSES OF THE TARGET WITH DIFFERENT APPARENT POSITIONS AND ILLUMINATION CONDITIONS

Exp.	Group	*N*	Light Source (Fig. 1)	*Ss'* Responses	
				Trapezoid Brighter	Square Brighter
I	A	6	*a*	5	0
	B	6	*a*	6	0
II	C	15	*a*	14	0
			b_1	0	14
III	D	5	*a*	5	0
	E	5	b_1	0	5
	F	5	*c*	0	0

whether a change in the apparent position of a target surface relative to an illumination source results in a change in the perceived brightness of that surface (cf. **7**, pp. 600, 612), even though the actual illumination conditions remain constant.

Experiment I

The apparatus[1] is shown in Fig. 1: the main illumination (100 w.) came from above (*a*), this being "indicated" to *S*s by the shadow distribution on several cubes (*d*). These "cues" are of considerable importance since, in preliminary experiments, little or no success was achieved without them. The *S*s looked monocularly through a reduction screen (*Sc*) at an upright cardboard trapezoid target (*t*), covered with Number 8 Hering gray paper, cut so that its retinal image would be the same as that of a square (*Sq*) lying flat on the black cloth surface (*F*). All *S*s ($N = 13$) reported seeing a horizontal square, whose brightness they were then (Judgment I) asked to match quickly and unanalytically to a scale (*H*) of Hering gray paper patches (Nos. 1, 3, 5, . . . 19). A round rod, *r*, ¾ in. in diameter, 22 in. long, painted white for Group A and black for Group B, was then waved behind the target (trying to avoid any cast shadows visible to *S*). This was kept up for some seconds, as it was difficult not to "see" the horizontal square (*Sq*) instead of the upright trapezoid (*t*); indeed, one *S* was dropped at this point, unable to see the target as upright. The *S*s again compared the target's brightness with the gray scale (Judgment II).

The results (Table 1) indicate that the target when apparently upright is reported as brighter than when apparently horizontal. Two questions may, however, be asked: First, while illumination of the rod is not likely to have been responsible for the brightness change since it was black for some *S*s and white for others, might not inadvertently cast shadows, or even the motion itself, have been the important factor? Second, if a stimulus is perceived as parallel to the line of regard, it should have a greater apparent area than when perceived as perpendicular to the line of regard (Fig. 1); may not the lower reported brightness in the former case be due to the smaller amount of *retinal* illumination per unit of *perceived* surface? The next two experiments were undertaken to test the first question by varying the means whereby the apparent shift in the target position is brought about, and to test the second by changing the direction of illumination while, of course, holding the apparent size change constant.

[1] Modified from one devised by Professor J. J. Gibson to study the relationship between perceived slant and perceived form.

Experiment II

The procedure of Exp. I was modified here in three ways: (*a*) When attempting to make the target appear upright, it was moved through short horizontal arcs (*R*) instead of having a rod (*r*) waved behind it. (*b*) The 15 *S*s (Group C) of this experiment ran through the procedure with the illumination coming from above (*a*), and then repeated the experiment with the illumination coming horizontally from in front of the trapezoid (b_1, Fig. 1), with a concealed supplementary source (b_2) to remove the shadow of the target (*t*) from the cues. (*c*) The *S*s were alternated in each part of the experiment as to the condition to which they were first subjected, the upright target or the horizontal square.

The results (Table 1) under illumination from above are the same as those in Exp. I: when seen upright, the target appears brighter than when seen flat. Under illumination from in front (b_1), the results are the reverse: with horizontal illumination, the target appears less bright when seen as upright than when flat. Since the change in perceived area consequent upon the change in perceived target position would be the same both when illumination comes from above and from the front, we can reject the amount-of-illumination-per-perceived-surface-area as a determining factor. This suggests that the brightness changes are due either largely or solely to the perceived change in target position with respect to the direction of illumination; this is supported by the results of the next experiment.

Experiment III

Two changes were here made in the procedures of Exp. I: (*a*) The target was changed

from flat to upright appearance by shifting from monocular vision to binocular vision (through a larger hole, *L*). (*b*) Illumination was maintained from above (*a*) as in Exp. I, for Group D ($N = 5$), from in front horizontally (b_1) as in the second part of Exp. II, for Group E ($N = 5$), and horizontally from the side (*c*), for Group F ($N = 5$).

The results of Groups D and E show the same changes in brightness as were found in Exp. II. There is no evidence of any perceived brightness changes in Group F, in which there is no change in perceived orientation of the target with respect to the illumination, since the illumination is parallel to the target surface in either of the two perceived positions. These results again suggest that the brightness changes are obtained due to the change in the relationship of the perceived direction of illumination and the perceived position of the surface it falls upon.

Discussion

In general, the results of these experiments suggest that when a surface of a given illumination is perceived as being perpendicular to the direction of illumination, it appears less bright than when the same surface, with the same illumination, seems parallel to the direction of illumination. How does this fit the various approaches to brightness constancy?

A simple one-to-one correspondence of illumination and perceived brightness must as usual be rejected, since the same stimulus arouses different responses. Likewise, any attempt to bring the perception of brightness into one-to-one correspondence with illumination *ratios* is inadequate, since differing responses are here obtained with the same illumination relationships. Either we must view Wallach's ratio formulation (or, for that matter, Helson's "adaptation level" explanation) as incomplete, or hold that there are at least two different kinds of brightness constancy, one bound to the illumination conditions and the other not, an unparsimonious position. The general viewpoint may, however, be retained (as may also a Gestalt organizational one) if the determinants of perceived brightness include not only the peripheral illumination relationships but the "cues" to spatial position and the illumination direction (cf. 7, p. 612). The empiricist or "inferential" position, disconcertingly enough, seems well able to explain the findings, at least by hindsight: thus, to reflect a given amount of light to the eye, a surface parallel to the incident illumination would have to have a higher albedo or brighter object color than would a surface perpendicular to the incident illumination, and would therefore be "inferred" to be brighter.

Summary

In order to determine whether perceived brightnesses can be brought into one-to-one correspondence with stimulus illumination *relationships* any more than with absolute illumination intensities, *S*s made judgments of the brightness of a target which, under constant or controlled conditions of illumination, was made to appear to be either perpendicular or parallel to the apparent direction of illumination. Since substantially the same illumination distributions produced different perceived brightnesses, analyses of brightness constancy in terms of stimulus illuminations cannot at present be considered complete explanations.

References

1. Brunswik, E. Über Farben-, Grossen- und Gestaltkonstanz in der Jugend. In H. Volkelt (Ed.), *Bericht über den XI. Kongress exp. Psychol.* Jena: Fischer, 1930. Pp. 52–56.
2. Gibson, J. J. *The perception of the visual world.* Cambridge: Houghton Mifflin, 1950.
3. Gogel, W. C., & Hess, E. H. A study of color constancy in the newly hatched chick by means of an innate color preference. *Amer. Psychologist*, 1951, **6**, 282. (Abstract)
4. Helson, H. Adaptation-level as frame of reference for prediction of psychological data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
5. Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.
6. Wallach, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, **38**, 310–324.
7. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received July 10, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

THE EFFECTS OF EXPERIMENTALLY INDUCED ATTITUDES UPON TASK PROFICIENCY

R. B. PAYNE AND G. T. HAUTY

USAF School of Aviation Medicine

It is conventional to suppose that subjective dispositions or biases can affect the course of many kinds of behavior, presumably through their directional and/or dynamic properties. This notion is especially prevalent in work-decrement literature, which frequently alleges that such factors as feelings, emotional tones, and task attitudes help determine task proficiency. In fact, it is central to those concepts of psychological fatigue which regard work decrement as symptomatic of cumulative task aversion and failure to maintain proper attitudes (**4**, **9**). Typically, however, such factors are merely inferred as intervening variables to account for differences in task proficiency. Seldom have investigators deliberately and knowingly induced differential dispositions toward a task in order to observe their effects upon output. Nevertheless, only by this means can the role played by such variables be finally understood.

The present study has sought to induce differential task dispositions by instructional and pharmacological means in order to observe their effects upon the maintenance of proficiency in a complicated compensatory pursuit task. While the results provide a reasonably lucid answer to the central question, they have the additional merit of depicting the nature of certain drug effects in independence of attitudes, and they reveal something about the combinatorial properties of dispositions generated by different methods.

Procedure

Subjects.—The *S*s were 80 volunteer, experimentally naive, basic airmen whose ages ranged narrowly around 20 yr.

Task.—Preliminary work indicated that the USAF SAM Multidimensional Pursuit Test (CM813E), developed by Loucks and described by Melton (**8**), would provide a task appropriate to the present study. Performance measures obtained thereon are typically quite reliable (Spearman–Brown corrected $r_{oo} = .85$ for $N = 1{,}800$), and, of greater importance, work decrement becomes manifest after about an hour's practice. In this task *S* is confronted with a panel containing four instruments, the pointers of which are made to drift eccentrically and unrelatedly about their respective nulls. Each pointer is associated electronically with a simulated aircraft control device (throttle, rudder, or stick) so that the drifts can be counteracted. Each instrument system includes its own scoring relay, which is energized when the pointer is within the tolerance limits of the null position. The secondary contacts of the relays are serially connected with the clutch of an automatic timer which sums and stores the time over which the four pointers have been kept in their null positions concurrently. The entire mechanism is under the control of a cycling system which meters out alternate work periods (1 min.) and rest periods (15 sec.) over any desired span of time.

In the present study the task was broken into two series of trials. Following the delivery of formal instructions by *E*, *S*s were given 40 trials in order to get beyond the early learning increments to a substantial level of skill. This was followed by a 10-min. interval during which they were assigned in a predetermined random order to experimental treatments soon to be described. After the treatments had been delivered, *S*s were put back to work on the task for a series of 192 trials (4 hr.).

Two copies of the test apparatus under the common control of a cycling device were used to test *S*s in pairs. No *S* could observe another's performance, nor did he receive information about his own progress aside from that gained momentarily from the pointers themselves. Instruments and controls were well lighted, ambient

noise was of low intensity, and temperature was held constant.

Experimental treatments.—Differential task dispositions were induced in two ways, (*a*) by instructional methods, and (*b*) by pharmacological methods. The instructional method was based upon the hypothesis that task performance is some positive function of the extent to which *S* is made to identify himself with the task and to appreciate the purpose and significance of what he is about to do. By this method we attempted to create two levels of attitude. To half of the *Ss* the following instruction, identified hereafter as I_1, was given:

"Now that you have learned to operate the apparatus we want to find out how well you can perform this task throughout a period of four hours. You are encouraged to do the very best you can during this time, for you will be giving us information that will help us deal with important practical problems."

The second instruction, identified hereafter as I_2, was given to the remaining half of the *Ss*. I_2 contained I_1 in principle, but it elaborated the theme at length, stressing the similarity between the task and real flight, the necessity for alertness, and the important role played by *S* in helping the Air Force solve one of its most important problems. Both instructions were delivered in a uniform way by one *E*.

The second method of induction took advantage of the analeptic and depressant properties of certain drugs (**4, 6, 10, 11, 12, 15, 16**). Each instructional group was divided into the following five treatment groups: (A) Control (no drug at all); (B) Placebo (lactose); (C) Analeptic 1 (5 mg. dextro-amphetamine sulfate); (D) Analeptic 2 (20 mg. caffeine derivative); (E) Depressant (0.65 mg. hyoscine hydrobromide mixed with 50 mg. diphenhydramine hydrochloride). All preparations were delivered in No. 1 pink capsules during the 10-min. interval between the two series of trials after delivery of the instructional treatments. They were introduced with the following statement:

"We will give you a capsule, the effects of which we are interested in studying. The only information we can give you is that the capsule is not the least bit harmful or injurious in any way. Are you willing to participate by taking this capsule? You may refuse and be excused now if you wish."

It will be apparent from the foregoing information that a 5×2 factorial design was employed, in which performance measures were taken both before and after the delivery of treatments. The 80 *Ss* were distributed equally and randomly among the subclasses.

Questionnaire.—At the conclusion of the work period, *Ss* completed a questionnaire designed primarily to elicit information about their covert dispositions as they reviewed the experience from which they had just emerged. No claim is made, of course, for any particular validity of the items for this purpose, nor for any particular reliability. Their outcomes in relation to the experimental treatments will help decide whether such properties can be inferred. The response choices for each item were fitted to a simple three-point scale, the value 3 being assigned in advance to the choice thought to indicate the most favorable task disposition, and the value 1, to the other extreme. The "disposition" items, choices, and assigned choice values were as follows:

1. Would you volunteer to go through something similar to this (*a*) tomorrow (3), (*b*) next week (2), (*c*) never (1); 2. How tired do you feel? (*a*) could work longer (3), (*b*) had enough (2), (*c*) ready to quit earlier (1); 3. Did you get bored or lose interest in what you were doing? (*a*) no or somewhat (3), (*b*) moderately (2), (*c*) very much (1); 4. What effect did the pill have upon your performance? (*a*) helped it (3), (*b*) no effect (2), (*c*) hindered it (1); 5. Do you think this experiment is of use to the Air Force? (*a*) much (3), (*b*) moderate (2), (*c*) little (1). Other items of a more specific and concrete nature required *Ss* to specify one of three degrees of fatigue for each of eight body locations.

TABLE 1

ANALYSIS OF VARIANCE OF ITEM RESPONSE VALUES

Source	*df*	Mean Square	*F*
Drugs	4	4.86	7.03
Motivations	1	2.56	3.70
Items	4	2.8537	8.49
D × M	4	.61	
D × I	16	.6834	2.03
M × I	4	.1287	
D × M × I	16	.3896	1.16
Between *Ss* treated alike	70	.6914	
Pooled interaction *Ss* × items	280	.336	
Total	399		

RESULTS AND DISCUSSION

Subjective effects.—A preliminary question is whether the experimental treatments did in fact induce the desired differences in task attitudes, as operationally defined. An analysis of the variance of item response values

will help decide. This analysis, shown in Table 1, is unnecessarily detailed for the problem at hand, but it serves our purpose well by showing that (*a*) the variation attributable to drugs is highly significant ($p < .001$), (*b*) the variation attributable to instructional groups is reasonably significant (p *circa* .03 when the null hypothesis is tested against the alternative $\bar{x}\ [I_2] > \bar{x}\ [I_1]$), and (*c*) the interaction of main effects is negligible ($F < 1$). In order to achieve a more definitive analysis of the drug variation, mean item response values were computed (see Table 2), and t tests of the null hypothesis against appropriate one-sided alternatives were performed. These show that the attitude level associated with the dextro-amphetamine treatment was superior to those associated with any other pharmacological variable (against the alternative C>A, $p = .001$; C>B, $p < .001$; C>D, $p = .04$; and C>E, $p = .001$). We can be only moderately confident of the caffeine derivative effects, despite the fact that this mean ranked in second place (D>A, $p = .12$; D>B, $p < .001$; and D>E, $p = .12$). Curiously enough, the mean of the depressant drug is exceeded only by the analeptic means. It is about equal to the control mean, but higher than the lactose mean ($p = .03$). The control mean is also higher than the lactose mean ($p = .03$).

In summary, we seem to have demonstrated the beneficial effects of an instructional technique (I_2) and two analeptic drugs upon attitudes, but not the negative effects of a depressant drug. No simple explanation capable of embracing all the facts seems readily available to account for the latter outcome. However, by assuming that the variation between control, placebo, and depressant treat-

TABLE 2

MEAN ITEM RESPONSE VALUES

Drug	Motive	Items					Means
		1	2	3	4	5	
A	I_1	2.13	2.75	2.63	2.00	2.50	2.40
	I_2	2.25	2.50	2.38	2.00	2.63	2.35
	Means	2.19	2.63	2.50	2.00	2.56	2.38
B	I_1	1.88	1.88	2.50	1.88	1.75	1.98
	I_2	2.38	1.88	2.88	1.50	2.50	2.23
	Means	2.13	1.88	2.69	1.69	2.13	2.10
C	I_1	2.63	2.88	3.00	2.63	2.75	2.78
	I_2	2.75	2.75	2.88	2.75	2.75	2.78
	Means	2.69	2.81	2.94	2.69	2.75	2.78
D	I_1	2.13	2.38	2.38	2.50	2.38	2.35
	I_2	2.25	2.75	2.88	2.75	2.88	2.70
	Means	2.19	2.56	2.63	2.63	2.63	2.53
E	I_1	2.38	2.50	2.38	1.50	2.50	2.25
	I_2	2.50	2.63	2.88	2.25	2.25	2.50
	Means	2.44	2.56	2.63	1.88	2.38	2.38
Column Means	I_1	2.23	2.48	2.58	2.10	2.38	2.35
	I_2	2.43	2.50	2.78	2.25	2.60	2.51
	Means	2.33	2.49	2.68	2.18	2.49	2.43

TABLE 3

ANALYSIS OF VARIANCE OF FATIGUE RATINGS

Source	*df*	Mean Square	*F*
Drugs	4	.3109	
Motivations	1	1.1391	1.66
Loci	7	21.3372	59.15
D × M	4	.7953	1.16
D × L	28	.3975	1.10
M × L	7	.3104	
D × M × L	28	.3908	1.08
Between *Ss* treated alike	70	.6872	
Pooled interaction *Ss* × loci	490	.3607	
Total	639		

ments is in fact fortuitous ($p \geqq .03$ in all comparisons), it is possible to view the analeptics merely as preservatives of initial attitude levels, which, in the absence of this preservative, would have dropped to minimum measurable levels. That we are in fact dealing here with diffuse, covert, task-oriented dispositions rather than realistically perceptible focal discomforts is supported by an analysis of *Ss*' ratings of fatigue intensity for each of eight bodily locations. As shown in Table 3, the only significant variation was that between body locations ($p < .001$), experimental treatments to the contrary notwithstanding. These locations were ordered in the following way with respect to rated fatigue intensity: arms (least), hands, neck, feet, eyes, legs, back, and buttocks (most), an order which seems intuitively valid, considering the nature of the task. Thus, it would seem that the experimental treatments established and maintained differential task dispositions despite the fact that *Ss* uniformly became tired in different places by significantly different amounts. This will be an important fact to recall when performance effects are appraised.

Performance effects.—The time-on scores made during the 40 pretreatment trials and the 192 posttreatment trials constitute the basic data for the statistical appraisal of performance effects. In order to eliminate score fluctuations due wholly to the eccentricity of the 8-min. cam pattern in the cycling system, the scores for each cycle of eight trials were summed for each *S*, thereby producing 5 pretreatment scores and 24 posttreatment scores for each *S*. Attention was then focused upon the final five cycles as the most rigorous test of the hypothesis under investigation. The mean performance levels exhibited during this period were converted to logarithms in order to homogeneate the subclass variances, and their regression upon pretreatment achievement levels was calculated. Following adjustment for regression, the treatment variances were tested against the estimate of error in order to determine their statistical significance. The analysis of adjusted variances is shown in Table 4, from which it is clear that very substantial differences were produced by the pharmacological treatments. By contrast, the difference between the instructional treatments was negligible ($F < 1$), and no appreciable significance could be attached to the interaction of the main effects.

A more definitive view of the pharmacological variation was obtained by

TABLE 4

ANALYSIS OF ADJUSTED VARIANCE OF TERMINAL PERFORMANCE

Source	*df*	Mean Square	*F*
Drugs	4	1.6648	33.30**
Motivations	1	.00004	
D × M	4	.0584	1.17
Error	69*	.0500	
Total	78*		

* Reduced by 1 *df* for regression ($r_{xy} = .58$; $b = .6862$).

** $p < .001$

TABLE 5

ADJUSTED TERMINAL PERFORMANCE MEANS

Drugs	Motivations		Means
	I_1	I_2	
A	1.9923	2.0252	2.0088
B	2.0188	1.9153	1.9671
C	2.3496	2.3427	2.3462
D	2.2488	2.1339	2.1914
E	1.4002	1.5828	1.4915
Means	2.0019	2.0000	2.0010

computing the adjusted means and testing their differences by the t statistic. The means are shown in Table 5. Tests of the null hypothesis against appropriate one-sided alternatives permit its rejection for all comparisons except that between control and lactose groups ($p \leqq .03$). This outcome can be more fully appreciated by reference to Fig. 1, which plots the general course of performance throughout the posttreatment period for each pharmacological group.

The broad picture is now partially clear. The instructional treatments induced significantly divergent task dispositions, as operationally defined, but their assumed dynamic properties were not evident in performance, nor did comparable analyses of earlier cycles reveal them. The analeptic treatments likewise allayed the development of unfavorable dispositions, but, as expected, they also allayed the loss of task proficiency. The question now is whether the performance outcome can be accounted for by the effects of drugs upon dispositions.

In view of the instructional outcomes, we shall consider the hypothesis about the effects of attitudes upon performance to be impugned if significant drug variation remains after the terminal performance scores have been adjusted for their multiple regression upon initial scores and attitude levels.

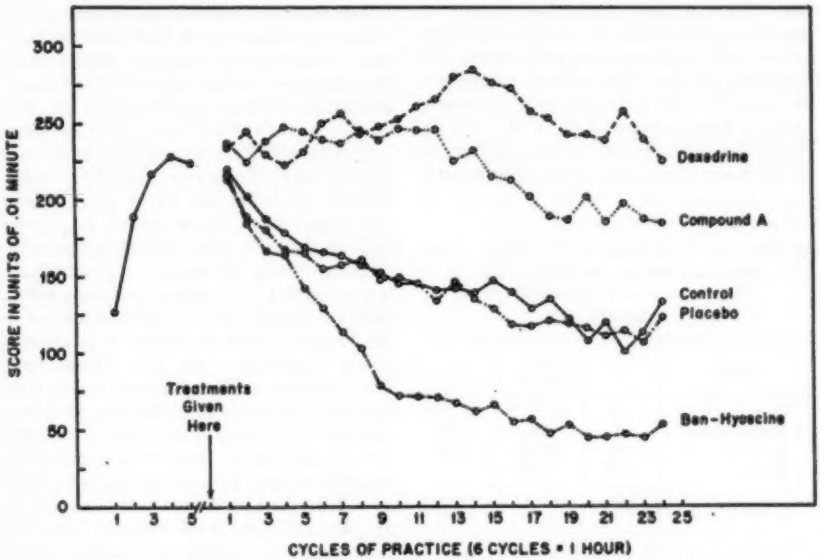

FIG. 1. Effects of the pharmacological treatments upon task performance

The analysis of the adjusted variances is given in Table 6, and the adjusted means are recorded in Table 7. These tables make it clear that very significant drug variation occurs in independence of the attitude variation. There is some decrement in the mean for dextro-amphetamine, but it remains significantly superior to the others. It is interesting to note in Table 6 that attitude scores are uncorrelated with pretreatment achievement scores ($r = .01$), but correlated somewhat with terminal scores ($r = .38$, $p < .01$). This constitutes some small bias in favor of the working hypothesis, but in the absence of confirming evidence from the instructional treatments it seems best to say that these covariant outcomes offer little if any indication that task dispositions constitute a link between the experimental treatments and terminal performance.

A final word should be said about the drug effects themselves. These results, although striking, are not novel (**4, 6, 15, 16**). They do, however, raise a question concerning the way in which the drugs achieve their effects upon performance, if not through the motivational properties of the subjective dispositions they create, as some have suggested (**1, 2, 6**). While this is largely a matter of speculation at the present time, it seems nevertheless

TABLE 6

Analysis of Variance of Terminal Performance after Adjustment for Attitude and Pretreatment Levels

Source	*df*	Mean Square	*F*
Drugs	4	1.3522	33.81**
Motivations	1	.0435	1.09
D × M	4	.0790	1.98
Error	68*	.0400	
Total	77*		

* Reduced by 2 *df* for multiple regression. $R_{y \cdot xz} = .69$; $r_{xz} = .01$; $r_{xy} = .38$; $r_{zy} = .58$.
** $p < .001$.

TABLE 7

Terminal Performance Means Adjusted for Attitude and Pretreatment Levels

Drugs	Motivations		Means
	I_1	I_2	
A	2.0012	2.0466	2.0239
B	2.1445	1.9717	2.0581
C	2.2536	2.2479	2.2508
D	2.2715	2.0596	2.1656
E	1.4499	1.5630	1.5065
Means	2.0241	1.9778	2.0010

likely that their effects can be viewed in terms of their presence in the body fluids which bathe or penetrate the nerve cells mediating task performance (**13, 14**). Their effects upon cell respiration, with implied changes in the efficiency of excitation and conduction processes, are probably of crucial importance. In order to view the effects in this way, it seems necessary first to postulate some sort of change in oxidative metabolism through repetitive exercise of the fibers mediating task performance (see **5**), a postulate which is accredited to some extent at the behavioral level by the fact that the positive influence of analeptic drugs upon performance is best demonstrated under conditions of task prolongation (**3, 17**). In the case of the amphetamines, depending upon the dosage, there is some evidence that enzymatic activity, carbohydrate metabolism, basal metabolism, and blood flow are increased, and some believe that the most logical explanation of their cellular effects lies in their capacity to inhibit the enzyme, amine oxidase, which normally alters certain enzymes in the brain to form aldehydes that depress tissue respiration (**6, 7**). Thus, they may enhance cell respiration, hence prolong the efficiency of excitation and conduction functions, by allaying the formation of aldehydes. Similar general systemic effects have been observed in the case of caffeine and its derivatives, but details at the cellular level remain obscure. Likewise, there is little information about the mechanism of action at the cellular

level in the case of the depressant preparation used.

Summary

Eighty *Ss* received preliminary training on a complicated compensatory pursuit task involving simulated aircraft instruments and controls, then continued work for 4 hr. after having been distributed among ten combinations of two motivational and five pharmacological conditions. The treatment effects were appraised in terms of subjective dispositions and task performance, and the functional connection between these outcomes was explored. Both motivational and pharmacological treatments induced divergent task dispositions. Motivational effects, however, were not discernible in performance. Large performance effects due to drugs could only slightly, if at all, be accounted for by subjective effects. While it may be conservative to suppose that these results are specific to the conditions of the present study, they nevertheless tend to discredit the general practice of invoking attitudinal constructs to explain work-decrement phenomena in the absence of confirming experimental demonstrations.

References

1. Barmack, J. E. The effect of benzedrine sulfate (benzyl methyl carbinamine) upon the report of boredom and other factors. *J. Psychol.*, 1938, **5**, 125–133.
2. Barmack, J. E. The effect of 10 mg. of benzedrine sulfate on the Otis test scores of college students. *Amer. J. Psychiat.*, 1940, **97**, 163–166.
3. Barmack, J. E. The time of administration and some effects of 2 grs. of alkaloid caffeine. *J. exp. Psychol.*, 1940, **27**, 690–698.
4. Bartley, S. H., & Chute, E. *Fatigue and impairment in man.* New York: McGraw-Hill, 1947.
5. Brink, F., Jr. Excitation and conduction in the neuron. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 50–93.
6. Goodman, L., & Gilman, A. *The pharmacological basis of therapeutics.* New York: Macmillan, 1941.
7. Krantz, J. C., & Carr, C. J. *The pharmacologic principles of medical practice.* Baltimore: Williams & Wilkins, 1951.
8. Melton, A. W. (Ed.) Apparatus tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No 4. Restricted Supplement.)
9. Myers, C. S. The study of fatigue. *J. Pers. Res.*, 1925, **3**, 321–334.
10. Payne, R. B., & Hauty, G. T. The effects of motion sickness preventives upon certain perceptual-motor components of the pilot's task. *USAF Sch. Aviat. Med., Proj.* No. 21-1601-0004, *Rep.* No. 3.
11. Payne, R. B., Moore, E. W., & Bethurum, J. L. The effects of certain motion sickness preventives upon psychological efficiency. *USAF Sch. Aviat. Med., Proj.* No. 21-32-019, *Rep.* No. 1.
12. Payne, R. B., Osier, D. R., & Tomlinson, P. A. S. The effects of certain motion sickness preventives upon navigator proficiency. *USAF Sch. Aviat. Med., Proj.* No. 21-1601-0004, *Rep.* No. 1.
13. Robson, J. M., & Keele, C. A. *Recent advances in pharmacology.* Philadelphia: Blakiston, 1950.
14. Salter, W. T. *A textbook of pharmacology.* Philadelphia: Saunders, 1952.
15. Shock, N. W. Some psychophysiological relations. *Psychol. Bull.*, 1939, **36**, 447–476.
16. Spragg, S. D. S. The effects of certain drugs on mental and motor efficiency. *Psychol. Bull.*, 1941, **38**, 354–363.
17. Tyler, D. B. The effect of amphetamine sulfate and some barbiturates on the fatigue produced by prolonged wakefulness. *Amer. J. Physiol.*, 1947, **150**, 253–262.

(Received July 23, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

TRANSFER OF PERCEPTUAL SET[1]

GORDON A. ECKSTRAND

Psychology Branch, Aero Medical Laboratory

AND DELOS D. WICKENS

The Ohio State University

Much of the research in the field of transfer of training has been concerned with the manner in which transfer varies as a function of the degree of similarity between the stimulus-response units of which the two tasks are composed, or between the individual stimuli or responses of the tasks. Frequently the concepts of stimulus and response generalization have been employed in theoretical analysis, and by so doing the analysis is freed to some extent of a dependence upon identical elements between the two tasks. Transfer then is couched in terms of stimulus dimension–response dimension units rather than specific S–R units.

There are, however, instances of transfer which cannot readily be understood in terms of the relationships among stimulus dimension and response dimension units. One example of this sort is found in the behavior which gave rise to Harlow's (1) concept of learning sets, for his monkeys showed a generalized ability to solve a *class* of problems even though the particular S–R associations acquired in one problem could not be transferred to the next. McGeoch and Irion (5) have used the term learning-how-to-learn to refer to certain laboratory experiments which demonstrate the same generalized kind of transfer, and other examples can be found in work on the solution of successive rational learning problems and in transfer in puzzle solution.

In fact, it is probable that in everyday life the most considerable portion of the significant positive transfer which occurs is between tasks for which no simple dimensional analysis of the stimulus and response similarity is possible. In other words, we learn how to solve problems much more frequently than solutions to problems.

In two recent experiments, Lawrence (2, 3) has presented data that suggest that rats may, through certain kinds of prior training, develop tendencies to disregard certain dimensions of sensory data, and this disregard may lead to negative transfer if in a later task this dimension contains the cue stimulus. It was the purpose of the present experiment to extend this type of transfer investigation into a more complicated problem with human *Ss*. Specifically, it is hypothesized that a training task functions to establish perceptual sets, and that, consequently, the utilization of a given cue or combination of cues in a later, similar task will be a function of the cue relevance pattern existing in the training task.

METHOD

Apparatus.—The apparatus consisted of a vertical display panel containing an aperture approximately at eye level and a horizontal reaction panel containing a starting key and four reaction keys. A 40-step stepping switch was mounted behind the display panel. By mounting an aluminum disc on the shaft of the stepping switch and placing 40 stimuli at proper intervals around the perimeter of the disc, it was possible

[1] The data for this experiment were collected at The Ohio State University under Contract AF 33(038)-15474 and have been previously reported in AF Technical Report No. 6566.

to present a continuous series of stimuli to *S*. Different series of stimuli could be presented simply by preparing other discs and changing them as desired.

The apparatus was wired so that there could be a fixed connection between a particular stimulus appearing in the aperture and a particular response key. The *S*'s task was that of learning to associate each of the four reaction keys with its appropriate stimulus. If the correct reaction key was depressed, a new stimulus appeared when *S* returned his hand to the starting key. If an incorrect key was depressed, no new stimulus appeared in the aperture when *S* returned to the starting key, and he had to continue in his attempts to select the correct key.

Procedure.—The stimuli used in this study were forms cut from construction paper of various colors. Thus, when *S* saw a stimulus in the aperture he could respond to its form, its color, or perhaps to both aspects of the stimulus. For each of the tasks used in the experiment, four forms and four colors were selected (Fig. 1). By arranging the forms on the stimulus disc in such a way that each of the four forms had a fixed connection with one of the response keys and by assigning colors to the forms in a random manner, it was possible to make form a relevant cue and color an irrelevant cue; i.e., the problem could be solved by utilizing the form quality of the stimulus, but not by utilizing its color quality. In a similar manner color could be made

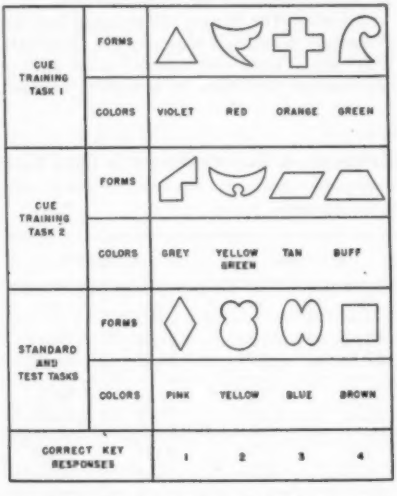

FIG. 1. Forms and colors used in the experimental tasks and correct key response for each stimulus

relevant and form irrelevant. By pairing color and form in a systematic manner and arranging the stimuli on the disc properly, it was possible to construct a task in which both color and form were relevant. Thus, it was possible to give *S*s three different kinds of cue training. They could be trained on (*a*) a task in which successful performance was possible only if color cues were attended to and form cues neglected, (*b*) a task which demanded that form cues be attended to and color cues neglected, or (*c*) a task which could be learned on the basis of either color or form or both.

Three groups of 40 *S*s each were used, one group for each of the three types of cue training. The cue training consisted of learning two separate tasks in which the forms and colors used were different, but the cue relevance pattern remained the same. This was done to accomplish more firmly the purpose of the training tasks in establishing cue attention habits and to minimize the tendency of *S*s to shift from using one cue to another merely as a function of going from one task to another. Figure 1 shows the forms and colors used in the two cue-training tasks as well as the particular key correct for each stimulus. Each of these tasks was learned to a criterion of ten consecutive errorless responses.

After completing the learning of their respective cue-training tasks, all three groups learned the same standard task. In this task, both form and color were relevant and the forms and colors used were different from those used in either of the cue-training tasks. Thus, the standard task was a completely new task in the sense that *S*s knew nothing about the stimulus-response associations to be formed. The three groups differed, however, in their previous experience with regard to what cues are relevant in such tasks. All groups learned the standard task to a criterion of ten consecutive errorless responses.

Following the preparatory and standard tasks, a test task was presented. This consisted of discs in which the stimulus objects were either colors alone or forms alone; they will be referred to hereafter as *test task-color* and *test task-form* respectively. These tasks were introduced to determine whether or not the preliminary training had a selective influence on which one of the cues available in the standard task was used to learn the task. Learning was again to a criterion of ten consecutive errorless responses. The form-relevant, color-absent condition was accomplished by cutting the appropriate forms from black construction paper. The color-relevant, form-absent condition was accomplished by having the appropriately colored construction paper cover enough of the disc so that color alone appeared in the aperture. The forms or the colors used in the two test tasks were the

TABLE 1

SCHEMATIC DESCRIPTION OF THE EXPERIMENTAL PROCEDURE

Task	Condition*					
	Consonant		Dissonant		Ambiguous	
	a	b	a	b	a	b
Training 1	F_rC_i	F_iC_r	F_iC_r	F_rC_i	F_rC_r	F_rC_r
Training 2	F_rC_i	F_iC_r	F_iC_r	F_rC_i	F_rC_r	F_rC_r
Standard	F_rC_r	F_rC_r	F_rC_r	F_rC_r	F_rC_r	F_rC_r
Test	F_rC_a	F_aC_r	F_rC_a	F_aC_r	F_rC_a	F_aC_r

* F = form, C = color, r = relevant, i = irrelevant, a = absent.

same as the forms and colors employed in the standard task. In addition, the S–R relationship between the forms or colors in the two test tasks was identical to the S–R relationship for the forms and colors in the standard task.

The rationale for utilizing this procedure is as follows. On the standard task all *S*s in all groups could have solved the problem by using either the form dimension, the color dimension, or both. If perceptual biasing operates, however, it would be expected that *S*s would be likely to be responsive to one dimension and not to the other, or at least less so. In addition, the dimension to which they are responsive would, we would expect, be determined by the previous experience of relevance-irrelevance during the two training tasks. One could not determine which dimension dominated in the standard task from an analysis of this task alone because both were available. In the test task only one dimension is available to *S*s. If this were the same dimension that had been employed in the standard task, then *S*'s performance on the test task should be better than if he had employed the other dimension during the learning of the standard task. For these *S*s, the test task constitutes essentially a recall task. Thus, by varying the dimension available in the test task one may infer which dimension was utilized in learning the standard task. If there were no differences among the groups in the learning of the test tasks, then, of course, there would be no evidence for perceptual biasing.

In summary, all groups learned the standard task with both cues relevant, after having had different experience with regard to cue relevance in the cue-training tasks. Performance on the test tasks, in each case, was used to determine what effects the preliminary task experience had upon the utilization of cues in learning the standard task.

The experimental procedure is indicated schematically in Table 1. The three major groups of the experiment are referred to by the terms consonant, dissonant, and ambiguous. By consonant it is meant that the same stimulus dimension which was required for solution in the first two tasks was the only dimension available in the test task; by dissonant it is meant that the dimension on the test task differed from that employed on the first two; and by ambiguous it is meant that both dimensions were usable in the first two tasks. For each major grouping there is a counterbalancing for form and color.

Subjects.—The *S*s were 120 students from the elementary psychology classes at The Ohio State University. They were assigned to the various experimental conditions in a random manner.

TABLE 2

MEAN NUMBER OF RESPONSES TO AND INCLUDING CRITERION FOR EACH CONDITION ON EACH TASK

Task		Condition					
		Consonant		Dissonant		Ambiguous	
		a	b	a	b	a	b
Training 1	*M*	67.25	65.00	92.60	61.10	44.95	53.80
	SD	35.94	39.14	47.32	34.95	24.92	35.50
Training 2	*M*	32.10	44.60	54.20	33.65	43.15	30.35
	SD	20.23	39.02	33.81	25.69	27.77	11.56
Standard	*M*	24.75	20.30	20.75	23.30	30.45	23.60
	SD	12.25	8.07	8.18	11.56	14.61	10.85
Test	*M*	16.25	12.10	28.00	17.25	24.10	15.85
	SD	6.56	2.51	12.95	5.92	14.25	7.19

Male and female *Ss* were nearly equally distributed within each group.

Results

Table 2 presents the mean number of responses to and including the criterion for each of the groups on each task. A response is defined as the occurrence of a single S–R sequence whether correct or not. Since the criterion responses are included in the scores presented in Table 2, ten should be subtracted from each score in order to find the number of responses preceding the first of the sequence of criterion responses.

The data relevant to the hypothesis are those that allow comparisons of the number of trials to learn the test tasks among the various experimental groups, i.e., among groups having learned the standard task following different types of cue training. If differences among such groups exist, we have evidence for the fact that a training task can function so as to exert an influence upon which of the cues available in another task situation will be utilized by the trainee.

The last row of Table 2 presents the average number of responses required by each of the experimental subgroups to learn the two test tasks to a criterion of ten consecutive errorless responses.

It is apparent from an examination of this table that the most rapid learning of the test task was found in the consonant groups and the least rapid in the dissonant. Because of inequality of variance among the groups, the Mann-Whitney *U* test was employed to determine the significance of the differences (**4**). The consonant form group differs from the dissonant form at a p level of .001, and from the ambiguous form at a level of .01; while the ambiguous form differs from the dissonant at a level of .02. The consonant color group differs from the dissonant at a level of .001, and from the ambiguous at a level of .11; while the ambiguous differs from the dissonant at the .05 level. These are all single-tailed tests.

Discussion

It may be inferred from the performance on the test tasks that the prior experience with the relevance and irrelevance of certain dimensions on the first two tasks developed biases which influenced the predominating cue on the third task. In this task either dimension—that of form or of color—could be employed, but the results imply that *Ss* were more responsive to one dimension than another and the direction of this responsiveness was related to the nature of the pretraining.

A simple stimulus generalization interpretation of these results would demand that there be a high degree of similarity of the form (or color) with which a response was associated in Tasks 1 and 2, and the form (or color) with which that response was associated in Task 3. An examination of Fig. 1 will indicate that no obvious systematic similarities of stimulus-response units persisted throughout the tasks. It is suggested that in tasks of this nature *Ss* may respond not only by pressing a key but may verbalize, stating perhaps "color (form) is relevant," not "yellow (square) for this key." This ability to name an entire dimension, and obtain reinforcement for that response, frees *S* from strict dependence upon physical or semantic similarity. It seems altogether possible that the behavior of subhumans would follow much more closely the prediction made from a stimulus-response generalization analysis. It would also seem probable that human *Ss* working with unfamiliar dimensions to which they have learned no inclusive symbol would show the more limited type of transfer expected from stimulus similarity. Wickens (**6**) has recently attempted to sketch out the S–R concepts which might be utilized in explaining transfer for nonverbalized dimensions,

but relatively little systematic work has been devoted to analysis of transfer which may be dependent upon verbalization of stimulus dimensions.

Summary

This experiment was conducted to test the influence of perceptual biases produced by pretraining upon learning in a later situation.

The task was a paired-associate type activity in which pressing one of four keys was determined by a visual stimulus. The visual stimuli contained two dimensions, form and color. Each *S* learned four tasks—two pretraining tasks which varied from group to group, a common task, and a test task in which only one of the two stimulus dimensions was present. The common task was one in which both dimensions were relevant to solution, and the question posed by the experiment was whether or not the pretraining would influence the dimension employed in learning this task. The operation of this bias was inferred by the performance on the final task in which only a single dimension was available to *S*s.

For a consonant group only one dimension, either form or color, was relevant to solution of the two pretraining tasks, and this same dimension was the sole dimension present in the test task. For a dissonant group, the dimension which was relevant in the pretraining was absent in the test task. For an ambiguous group both dimensions were relevant in the pretraining tasks.

The influence of the pretraining was shown by the fact that on the test task the consonant group learned with greatest ease, the ambiguous was next, and the dissonant was the poorest.

References

1. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
2. Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
3. Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
4. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
5. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
6. Wickens, D. D. An interpretation of the development of a perceptual set in S–R terms. *Wright Air Development Center, USAF, Tech. Rep.* 52-305, September, 1952.

(Received July 9, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

JUDGMENT OF TIME AS A FUNCTION OF SERIAL POSITION AND STRESS

JOHN L. FALK[1] AND DALBIR BINDRA

McGill University

In a recent experiment on temporal judgment Eson and Kafka (**2**) noted that overestimation of time was greater in the initial trials (judgments) than in later trials of the same series. However, these investigators worked with only a short series of four trials, and did not systematically explore the effect of serial position on temporal judgment. One of the objects of the present experiment was to study this effect over a greater range of trials.

A second aim of this study was to investigate the effect of mild stress on the judgment of time. Postman and Schneider (**6**) have pointed out that for needs, expectancies, etc. to measurably affect perception, the "stimulus information" given to *S* must be meager or poorly structured. Thus, judgments of time, with the paucity of cues on which they are based, and their dependence upon memory, should provide an ideal means for studying the effects of directive or motivational factors in perception. A number of earlier studies have shown that specific motivational factors, such as goal-direction and "need tension" affect the estimation of time (**4, 5, 7**). In the present experiment *stress* was used as a motivational variable, and the judgments of time made while *S*s anticipated receiving an electric shock were compared with judgments made under nonstress conditions.

METHOD

Subjects.—Forty *S*s were divided equally between an experimental and a control group. The *S*s were mainly college students, with a mean age of 25 yr. and a range of 15 to 36 yr. Each group contained 14 males and 6 females.

Apparatus.—The apparatus consisted of the following items, powered by appropriate arrangements of line-current and dry-cell batteries: (*a*) a chronometer (1/100 sec.) was set in operation when *S* pressed a key, and stopped when *S* released it; (*b*) a memory drum which exposed the various signals to *S*; (*c*) an inductorium which delivered a mild shock to *S* via a pair of spring clamps over the fingers of his left hand (the shock was usually described as somewhat painful, and *S* usually winced and made slight withdrawal movements when it was administered); (*d*) a polygraph which produced a continuous record of palmar resistance; (*e*) a small desk fan, used to mask the sounds coming from *E*'s side of a partitioning screen; and (*f*) a standard doorbell wired to a push button.

The production method.—The production method of temporal judgment was used in the present study. In this method, *S* marks off an interval which has been verbally stated by *E*. In our experiment *S* was asked to mark off 15 sec. by pressing and then releasing a key. The actual time for which *S* keeps the key pressed will be called *elapsed time* (**2**), and it represents the length of time that, in *S*'s judgment, is equal to 15 sec. In using the production method, overestimation of time means that *S* allows less chronological time to elapse before he considers the stated value as having been reached. Thus, the smaller the elapsed time, the greater the overestimation.

Procedure.—The *S*s were tested individually. Each *S* was informed that the experiment would last about 45 min., and appointments were arranged only for those times at which *S*s had no pressing obligations. If, during the experiment, *S* asked how much longer the experiment would last, he was told: "I would rather not say, since it is possible that it may affect your results. However, you know that the whole experiment lasts about 45 min." Only two or three *S*s asked for such information. The *S*s were given no clue as to the accuracy of their time judgments.

In the experimental group the GSR electrodes, well smeared with electrode jelly, were attached to *S*'s right hand.[2] The index and

[1] This is a condensed presentation of material submitted in partial fulfillment of the requirements for the M.A. degree at McGill University (**3**). This research was supported in part by the National Research Council of Canada.

[2] Although GSR records were obtained from all *S*s, their analysis has not been carried out. As is well known, meaningful analysis of GSR

middle fingers of his left hand were placed in the spring clamps which were wired to the inductorium. After noting the basal resistance of *S*'s palm, *E* came from behind the partitioning screen and read the following instructions to *S*:

"This is an experiment in time estimation. You are to attempt to estimate an interval of 15 sec. on each trial. You will make your estimate of 15 sec. by pressing down this key (*E* indicates the chronometer key to *S*) and holding it down until you think 15 sec. have gone by. Then release it. Try pressing the key and holding it down, just to get the feel of it. (*S* presses and releases the key a few times.) Just before each trial I will say 'ready' and then as a signal for you to start your estimation you will hear a click from this box (*E* points to the memory drum), and you will see either a red or a blue mark appear in this slot (*E* points to the slot in the memory drum). Then you will press the key down and keep looking at the mark until you think 15 sec. have gone by. Then release the key.

"Try not to shift your position or move about much, especially during the trials. At other times tell me if you shift, move the chair, or yawn. Keep your right arm still, in the position most comfortable to you where you can reach the key easily. That's because this (*E* points to the GSR electrodes) is such a sensitive indicator of activity. When you finish each estimate, that is, when you take your finger off the key, you will either hear a bell ring or feel a slight shock in your two fingers. If the mark showing has been blue, you will hear a bell ring like this. (*E* rings the bell.) If the mark showing has been red, you will feel a slight shock in your fingers like this. (*E* shocks *S*.)

"Remember: First I will say 'ready.' Then comes a click which means you are to begin. Look at the mark which appears in the slot, press the key down, and hold it down until you think 15 sec. have passed. Do not try to count off the seconds to yourself. I'm just interested in what you *feel* 15 sec. is like—so don't count. When you feel that 15 sec. have passed, release the key. There will be a short rest between trials. Do you have any questions? (There were usually no questions.) All right. We'll try a few practice trials to begin with."

The sequence was as described in the instructions. The pause between trials (from the occurrence of the bell or shock to the beginning of the next trial) was about 25 to 30 sec.

Five "practice trials" were given during which only blue marks appeared, and consequently only bells (no shock) occurred when the key was released at the conclusion of the judgment. When five practice trials were completed *S* was told: "I think you've had enough practice now. You've got the idea. Now we'll begin." Each *S* was given 20 trials, 10 with blue (bell) signals and 10 with red (shock) signals in a predetermined "random" order.

When these trials were completed, *E* told *S:* "The experiment is actually over now, but I'd like to do just a few more trials for the sake of comparison. Is that all right?" The *S* invariably consented, and five more "extra trials" were given during which only blue signals appeared. Thus there were altogether 30 trials.

The control group was treated in exactly the same way as the experimental group, except that no mention of shock was made in the instructions, nor was any shock administered. This group was told that after a trial with a red signal they would "hear a slight twanging sound." The "twanging sound," which was present for the experimental group as well, was produced by pressing and subsequently releasing the button (which delivered the shock to the experimental group) on the inductorium.

Results

The 30 judgments (5 practice trials, 20 experimental trials, and 5 extra trials) made by each *S* were divided into six blocks of five judgments each, and the mean of each block of five successive judgments was calculated. These subject means were then averaged for the 20 *S*s. The six serial means obtained for each group are plotted in Fig. 1, and were found to lie in a more or less linear fashion for both the experimental and control

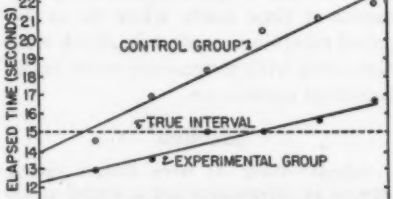

Fig. 1. Mean elapsed time of intervals judged to be 15 sec., as a function of the serial order of the judgment

data is fraught with theoretical as well as practical difficulties. However, from a cursory examination of the records, a hypothesis connecting palmar conductance to temporal judgments in any simple manner would appear to be ill-founded.

TABLE 1

DURATION (IN SECONDS) OF JUDGED 15-SEC. INTERVALS UNDER VARIOUS CONDITIONS

Measure	Experimental Group		Control Group	
	Mean	*SD*	Mean	*SD*
Blue signal (bell)	14.9	5.0	18.8	12.3
Red signal (shock for exp. group)	14.6	4.8	19.6	13.6
Total (including practice and extra trials)	14.8	5.4	18.9	13.4

groups. This relationship will hereafter be referred to as the *serial position gradient.*

Since both the serial position gradients approximated linearity, the usual methods for testing the significance of the slope of a line from a slope of zero, and for testing the significance of the difference between two slopes, were used. The regression coefficient, or slope of the line, is significantly different from zero for the experimental group ($p<.01$) and for the control group ($p<.05$), but the difference between the two regression coefficients is not significant ($p>.10$). The difference between the *SD*'s of the groups (see Table 1—Total) is significant ($p<.001$), and the difference between the means is significant ($p<.01$).

Breaking down the 20 "experimental trials," the means of the blue-signal (bell) trials and red-signal (shock for experimental group) trials were calculated. These means are not significantly different for either the experimental or the control group.

DISCUSSION

The serial position gradient.—The gradients for both groups confirm the one found by Eson and Kafka (2), who also used the production method. These *Es* demonstrated a similar, significant serial position gradient over the four trials they utilized, and concluded that there was, in general, overestimation of time in a time-estimation–oriented task, the effect of serial position being to successively decrease the amount of this overestimation. Eson and Kafka interpreted this trend as an *improvement* in judging the interval (15 sec.). Such an interpretation is not supported by our results, since with an increased number of trials the elapsed time (*S*'s judgment) passes through 15 sec. and continues to increase.

The work done on the time-order error in temporal judgments appears to rule it out as an explanation of the serial position gradient. Both Stott (8) and Woodrow (9) found that with the comparison of two equal intervals, the second one is *overestimated.* Therefore, the conclusion may be drawn that *if* a time-order error is operative under the conditions of the production method, the serial position gradient occurs in spite of it.

Effect of stress.—The experimental group was given shock and nonshock trials. It was felt that the expectancy of pain on the shock trials would induce a mild state of anxiety in *Ss* during the judging of those particular intervals. Under such uncomfortable circumstances, it was hypothesized that time would be overestimated. The situation would parallel the overestimation of time typical in states of pain or intense anticipation. However, no significant difference was found between the shock and the nonshock trials for the experimental group.

It is possible that *Ss* did not discriminate motivationally between the shock and nonshock signals, but that they, nevertheless, were affected by the shock situation. This appears to be a valid interpretation in view of the fact that the position of the gradient for the experimental group is displaced downward, i.e., toward overestimation, with respect to the control group, the mean elapsed time for the experimental group being significantly lower than that of the control group. Evidently, the knowledge that shock was forthcoming did not affect only the shock trials, but affected equally all the judgments made in the shock situation (experimental group).

The results also show that the difference between the two slopes is not statistically significant. This means that the administration of shock, as carried out in

this experiment, does not affect the *slope* of the serial position gradient.

The significant downward displacement of the experimental group's gradient may be interpreted as either (*a*) a shift toward greater accuracy, or (*b*) a shift due to the anticipation of painful shock.

Concerning the first alternative, an argument similar to the following could be constructed for its support. The shock, by making *S*s more anxious, may have served to make the experimental group, on the whole, more alert or "vigilant." With this higher level of "vigilance" operating, *S*s would tend to be more accurate; hence the closer approximation to 15 sec. achieved by the experimental group (mean = 14.8 sec.). Bruner and Postman (1) have proposed the construct of "vigilance" to explain cases in which increased stress is associated with increased accuracy of perception. Thus, the smaller deviation of the mean of the elapsed times from 15 sec. in the case of the experimental group could be interpreted as indicating a "vigilant" attitude resulting in greater over-all accuracy of judgment.

The second alternative makes no assumptions concerning the relatively greater accuracy of the experimental group. The finding that the experimental group's judgments are displaced toward overestimation can be looked upon as being independent of the notion of accuracy, the displacement of the judgments being attributed solely to the more anxious *set* induced by shock. Such a set also means that *E* has controlled the "mental content" of these *S*s to a greater extent than that of the control *S*s, thereby reducing the variability (*SD*). We prefer this second alternative: The smaller mean of the experimental group is not the result of increased accuracy due to "vigilance," but represents a general displacement of the judgments toward overestimation (relative to the control group) owing to a more anxious set induced by the expectation of shock.

Summary

The object of this study was to explore the effects of the serial position of judgment and mild stress, induced by the threat of impending shock, upon judgment of a 15-sec. interval by the production method. An experimental group ($N = 20$) received shock at the end of certain trials, and a control group ($N = 20$) never received shock. The *S* was forewarned at the beginning of each time interval of the type of consequences to be received at the end of the interval.

Statistically significant serial position gradients for temporal judgments were found in both groups. These gradients were more or less linear, and ran from overestimation in the earlier trials to underestimation in the later trials.

The anxiety-producing situation (experimental group) led to a significantly greater overestimation than a comparatively neutral situation, but the direction and slope of the serial position gradient was not significantly altered by the differences in treatment given the experimental and control groups.

It is argued that the serial position gradient does not result either from time-order error, or from an improvement in ability to judge time. Greater overestimation of the interval by the experimental group (relative to the control group) is interpreted in terms of an anxious *set* induced by the expectation of shock.

References

1. Bruner, J. S., & Postman, L. An approach to social perception. In W. Dennis (Ed.), *Current trends in social psychology*. Pittsburgh: Univ. of Pittsburgh Press, 1948. Pp. 71–118.
2. Eson, M. E., & Kafka, J. S. Diagnostic implications of a study in time perception. *J. gen. Psychol.*, 1952, **46**, 169–183.
3. Falk, J. L. Some factors affecting the judgment of temporal intervals. Unpublished M.A. thesis, McGill Univ., 1952.
4. Filer, R. J., & Meals, D. W. The effect of motivating conditions on the estimation of time. *J. exp. Psychol.*, 1949, **39**, 327–331.
5. Hindle, H. M. Time estimates as a function of distance traveled and relative clarity of a goal. *J. Pers.*, 1951, **19**, 483–501.
6. Postman, L., & Schneider, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, **58**, 271–284.
7. Rosenzweig, S., & Koht, A. C. The experience of duration as affected by need-tension. *J. exp. Psychol.*, 1933, **16**, 745–775.
8. Stott, L. H. Time-order errors in the discrimination of short tonal durations. *J. exp. Psychol.*, 1935, **18**, 741–766.
9. Woodrow, H. The effect of practice upon time-order errors in the comparison of temporal intervals. *Psychol. Rev.*, 1935, **42**, 127–152.

(Received June 30, 1953)

Journal of Experimental Psychology
Vol. 47, No. 4, 1954

SMOOTH EYE MOVEMENTS IN THE ABSENCE OF A MOVING VISUAL STIMULUS[1]

GERALD WESTHEIMER

University of Houston

AND DONALD W. CONOVER

U. S. Navy Electronics Laboratory

Conjugate eye movements involving cortical areas are generally classified (3) under two headings: (*a*) saccadic movements, and (*b*) pursuit movements. The two types of movements have quite different characteristics and are easily distinguishable on photographic eye movement records. Pursuit movements are generally very smooth and involve relatively slow changes in the position of the eye in the orbit. Conversely, saccadic movements of the eyes are rapid and show defined changes in velocity during their course (5).

It is generally stated that the stimulus to pursuit movements must be a moving stimulus and that pursuit movements cannot be executed in the absence of such a stimulus (6). Supporting evidence for this view is sought in pathological conditions in which patients with certain neurological lesions are said not to be able to move their eyes "at will" but can follow a slowly moving stimulus object (1). This type of evidence has led to the postulation of neurological pathways subserving eye movements (4) which separate the centers and some of the pathways into those subserving "voluntary movements" and those concerned with "involuntary" or "reflex" or "pursuit" movements. No clear distinction is made in the literature between "smooth" and pursuit-type movements, and it is generally stated that all eye movements (excepting those of labyrinthine or other lower reflex origin) are saccadic unless there is a moving visual stimulus in the field. The purpose of this paper is to report the occurrence of smooth eye movements in the absence of an external moving visual stimulus.

[1] Based on an experiment carried out in the Physiological Optics Laboratory of The Ohio State University.

METHOD

Apparatus.—The recording equipment used in this experiment consisted of a modified American Optical Company "Ophthalmograph" eye movement camera. The modifications were chiefly in the direction of increasing the resolution in the time dimension. By appropriate optical devices within the camera, only the horizontal components of the motions of the two eyes were recorded. For a precise time base on the film, a beam of light, interrupted 100 times a second by a synchronous chopper, was recorded along with the corneal reflex from both eyes. The *S*s were firmly anchored to the camera with a full impression of their teeth in a bite bar. The only visual stimuli in *S*'s field of regard, other than the lights used to produce the corneal reflection, were two fixation-point lights (10′ in diameter) separated by 30°, at the nodal point of the eye projected upon a large screen at a distance of 150 cm. from *S*'s eye.

Procedure.—The method is similar to those employed by Westheimer (5) in his study of eye movements in response to visual stimuli in the horizontal plane, and by Conover and Stroud (2) in their investigation of eye movements in following what they described as imaginary target motions. No special preparation is required of *S*; two students with normal ocular motility and without previous training in the experimental task were used as *S*s of this study.

Trial runs (using other *S*s) had shown that the experimental task could be carried out in a lighted room or in the dark without significant effects upon *S*'s performance. The two fixation

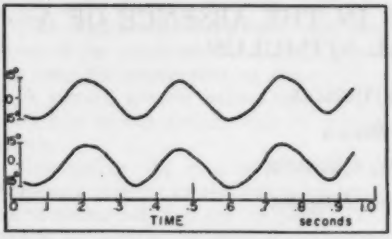

FIG. 1. Tracing of photographic record of movements of both eyes in response to instructions to move eyes back and forth smoothly and rapidly between two luminous points 30° apart in a horizontal plane

lights were turned on and *S* was instructed to move his eyes as rapidly and as smoothly as possible between the two fixation points. Fifteen seconds of *S*'s response were recorded each session, making a total of 45 sec. of record for each of the two *S*s.

RESULTS

In both *S*s considerable stretches of the eye movement records obtained in this situation looked like the excerpt shown in Fig. 1. The movements have all the characteristics of smooth or pursuit-type motions of the eye. For purposes of comparison, a section of a record containing both smooth movements and saccadic movements is shown in Fig. 2. The record shown in Fig. 1 is, in all respects, identical with records obtained of the eyes in following a stimulus moving in the horizontal direction with velocity and acceleration changes dictated by a sine function at the maximum rate at which these can be carried out, viz., about three cycles per second. When *S*s attempted to move their eyes smoothly but slowly from one position to the other, no such pursuit-type movement was evident.

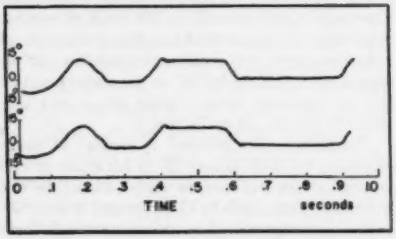

FIG. 2. Tracing of photographic record of movements of both eyes in response to instructions to move back and forth smoothly between two luminous points 30° apart in a horizontal plane. Tracing shows both smooth movements and saccadic movements.

SUMMARY

It was demonstrated that eye movements fitting in all respects the conventional description of smooth or pursuit movements can occur in the absence of a moving visual stimulus in a situation where there is no reason to expect a lower reflex arc to be in operation. It has generally been believed that such movements require for their initiation the presence of a moving visual stimulus. The evidence presented here suggests that, at least in the case of the present two *S*s, the pathways subserving smooth movements can be activated without a moving visual stimulus.

REFERENCES

1. COGAN, D. *The neurology of the ocular muscles.* Springfield, Ill.: Charles C Thomas, 1948.
2. CONOVER, D., & STROUD, J. Eye movements in following real, apparent, and imaginary target motions. Paper read before the IX Annual Research Conference on Vision, Ohio State Univ., June, 1951.
3. DUKE-ELDER, W. S. *Text-book of ophthalmology.* Vol. I. London: Henry Kimpton, 1942.
4. SPIEGEL, E. A., & SOMMER, I. *Neurology of the eye, ear, nose, and throat.* New York: Grune & Stratton, 1944.
5. WESTHEIMER, G. The response of the oculomotor system to visual stimuli in the horizontal plane. Unpublished Ph.D. dissertation, Ohio State Univ., 1953.
6. WOODWORTH, R. S. *Experimental psychology.* New York: Henry Holt, 1938.

(Received July 20, 1953)